(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4087097号
(P4087097)
(45)発行日　平成20年5月14日(2008.5.14)　(24)登録日　平成20年2月29日(2008.2.29)

(51)Int.Cl.　F I
G06F 12/00 (2006.01)　G06F 12/00 501B
G06F 17/30 (2006.01)　G06F 12/00 513Z
G06F 17/30 110C
G06F 17/30 240Z

請求項の数 14　(全 42 頁)

(21)出願番号　特願2001-345523(P2001-345523)
(22)出願日　平成13年11月12日(2001.11.12)
(65)公開番号　特開2003-150414(P2003-150414A)
(43)公開日　平成15年5月23日(2003.5.23)
審査請求日　平成15年12月24日(2003.12.24)

(73)特許権者　000005108
株式会社日立製作所
東京都千代田区丸の内一丁目６番６号
(74)代理人　100079108
弁理士　稲葉　良幸
(74)代理人　100093861
弁理士　大賀　眞司
(72)発明者　茂木　和彦
神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所　システム開発研究所内
(72)発明者　大枝　高
神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所　システム開発研究所内

最終頁に続く

(54)【発明の名称】データベース管理システム情報を考慮したデータ再配置方法およびデータ再配置を行う計算機システム

(57)【特許請求の範囲】
【請求項１】
データベース管理システムが稼動する少なくとも１台以上の計算機と、
少なくとも１つ以上のデータを記憶する物理記憶手段を有し、前記データベース管理システムにより管理されるデータベースデータを格納した少なくとも１台以上の記憶装置と、
前記計算機と前記記憶装置との間に接続され、前記計算機と前記記憶装置との間でデータの転送を制御する記憶制御手段と、
前記記憶装置における前記データの配置を管理するデータ位置管理サーバとを有する計算機システムにおけるデータの再配置方法において、
前記データ位置管理サーバが、前記計算機から前記データベース管理システムにより管理されるデータベースに関する情報、前記記憶制御手段からデータの記憶位置に関する情報及び前記記憶装置から前記物理記憶手段の稼動情報である物理記憶稼動情報を取得するステップと、
前記データ位置管理サーバが、前記データベースに関する情報、前記データの記憶位置に関する情報及び前記物理記憶稼動情報からなる取得情報に基づいて、前記データベースデータの配置を決定するステップと、
前記データ位置管理サーバから前記決定されたデータ配置を実現するためのデータ移動を前記記憶制御手段に指示するステップと、
前記記憶制御手段が、前記指示に従って、前記記憶装置内に格納された前記データの配

置を変更するステップとを有し、
前記データベースデータの配置を決定するステップでは、前記データの配置を決定する際に、前記取得情報に基づいて、同時にアクセスされる可能性が高い前記データベースデータの組を検出し、前記組を異なる前記物理記憶手段に配置するように、かつ、前記物理記憶手段における記憶位置まで定めたデータの配置を決定することを特徴とするデータの再配置方法。
【請求項２】
前記データ位置管理サーバの機能が、所定の前記データベースが稼動する計算機上で実現されることを特徴とする請求項１記載のデータの再配置方法。
【請求項３】
前記データ位置管理サーバの機能が、所定の前記記憶制御手段上で実現されることを特徴とする請求項１記載のデータの再配置方法。
【請求項４】
前記データ位置管理サーバの機能が、所定の前記記憶装置上で実現されることを特徴とする請求項１記載のデータの再配置方法。
【請求項５】
前記記憶制御手段が、前記計算機上で実施されるプログラムにより実現されることを特徴とする請求項１記載のデータの再配置方法。
【請求項６】
前記データ位置管理サーバから前記記憶制御手段に決定されたデータの配置を指示する前に、前記決定されたデータの配置を管理者に提示し、管理者にデータの配置変更を実施するか確認することを特徴とする請求項１記載のデータの再配置方法。
【請求項７】
前記データベースに関する情報は、前記データベース管理システムのスキーマにより定義される表・索引・ログを含むデータ構造に関する情報と、前記データベースデータを前記スキーマにより定義されるデータ構造毎に分類した前記記憶装置における記録位置に関する情報の少なくとも１つを含むことを特徴とする請求項１に記載のデータの再配置方法。
【請求項８】
データベース管理システムが稼動する少なくとも１台以上の計算機と、
少なくとも１つ以上のデータを記憶する物理記憶手段を有し、前記データベース管理システムにより管理されるデータベースデータを格納した少なくとも１台以上の記憶装置と、
前記計算機と前記記憶装置との間に接続され、前記計算機と前記記憶装置との間でデータの転送を制御する記憶制御手段と、
前記記憶装置における前記データの配置を管理するデータ位置管理サーバとを有する計算機システムにおいて、
前記データ位置管理サーバは、前記計算機から前記データベース管理システムにより管理されるデータベースに関する情報、前記記憶制御手段からデータの記憶位置に関する情報及び前記記憶装置から前記物理記憶手段の稼動情報である物理記憶稼動情報を取得する情報取得手段と、
前記データベースに関する情報、前記データの記憶位置に関する情報及び前記物理記憶稼動情報からなる取得情報に基づいて、前記データベースデータの配置を決定する配置決定手段と、
前記データ位置管理サーバから前記記憶制御手段により決定されたデータの配置を実現するためのデータ移動を指示するデータ配置指示手段を有し、
前記記憶制御手段は、前記データ位置管理サーバからの指示に従って、前記複数の記憶装置内に格納された前記データの配置を変更するデータ配置変更手段を有し、
前記配置決定手段は、前記データの配置を決定する際に、前記取得情報に基づいて、同時にアクセスされる可能性が高い前記データベースデータの組を検出し、前記組を異なる

前記物理記憶手段に配置するように、かつ、前記物理記憶手段における記憶位置まで定めたデータの配置を決定することを特徴とする計算機システム。
【請求項９】
前記データ位置管理サーバと所定の前記データベースが稼動する計算機が同一の計算機であることを特徴とする請求項８記載の計算機システム。
【請求項１０】
前記データ位置管理サーバと前記記憶制御手段が同一の装置であることを特徴とする請求項８記載の計算機システム。
【請求項１１】
前記データ位置管理サーバと所定の前記記憶装置が同一の装置であることを特徴とする請求項８記載の計算機システム。
【請求項１２】
前記記憶制御手段が、前記計算機上で実施されるプログラムにより実現されることを特徴とする請求項８記載の計算機システム。
【請求項１３】
前記配置決定手段により決定したデータの配置を管理者に提示する手段と、前記管理者からデータの配置変更の可否を取得する手段を有することを特徴とする請求項８記載の計算機システム。
【請求項１４】
前記データベースに関する情報は、前記データベース管理システムのスキーマにより定義される表・索引・ログを含むデータ構造に関する情報と、前記データベースデータを前記スキーマにより定義されるデータ構造毎に分類した前記記憶装置における記録位置に関する情報の少なくとも１つを含むことを特徴とする請求項８に記載の計算機システム。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、データベース管理システムに関する。
【０００２】
【従来の技術】
現在、データベース（ＤＢ）を基盤とする多くのアプリケーションが存在し、ＤＢに関する一連の処理・管理を行うソフトウェアであるデータベース管理システム（ＤＢＭＳ）は極めて重要なものとなっている。特に、ＤＢＭＳの処理性能はＤＢを利用するアプリケーションの性能も決定するため、ＤＢＭＳの処理性能の向上は極めて重要である。
【０００３】
ＤＢの特徴の１つは、多大なデータを扱うことである。そのため、ＤＢＭＳの実行環境の多くにおいては、ＤＢＭＳが実行される計算機に対して大容量の記憶装置あるいは複数の記憶装置を接続し、記憶装置上にＤＢのデータを記憶する。そのため、ＤＢに関する処理を行う場合に、記憶装置に対してアクセスが発生し、記憶装置におけるデータアクセス性能がＤＢＭＳの性能を大きく左右する。そのため、ＤＢＭＳが稼動するシステムにおいて、記憶装置の性能最適化が極めて重要であり、特にＤＢＭＳが管理するデータの物理記憶装置に対する配置の最適化が重要である。
文献“Ｏｒａｃｌｅ８ｉパフォーマンスのための設計およびチューニング，リリース８．１”（日本オラクル社，部品番号Ｊ００９２１－０１）の第２０章（文献１）は、ＲＤＢＭＳであるＯｒａｃｌｅ８ｉにおけるＩ／Ｏのチューニングを述べている。その中で、ＲＤＢＭＳの内部動作のチューニングと共に、データの配置のチューニングに関連するものとして、ログファイルは他のデータファイルから分離した物理記憶装置に記憶すること、ストライプ化されたディスクにデータを記憶することによる負荷分散が効果があること、表のデータとそれに対応する索引データは異なる物理記憶装置に記憶すると効果があること、ＲＤＢＭＳとは関係ないデータを異なる物理記憶装置に記憶することが述べられている。

【０００４】
米国特許６０３５３０６（文献２）は、ＤＢＭＳ－ファイルシステム－ボリュームマネージャ－記憶装置間のマッピングを考慮した性能解析ツールに関する技術を開示している。この性能解析ツールは、各レイヤにおけるオブジェクトの稼動状況を画面に表示する。このときに上記のマッピングを考慮し、その各オブジェクトに対応する他レイヤのオブジェクトの稼動状況を示す画面を容易に表示する機能を提供する。また、ボリュームマネージャレイヤのオブジェクトに関して、負荷が高い記憶装置群に記憶されているオブジェクトのうち、２番目に負荷が高いオブジェクトを、もっとも負荷が低い記憶装置群に移動するオブジェクト再配置案を作成する機能を有している。
【０００５】
特開平９－２７４５４４号公報（文献３）は、計算機がアクセスするために利用する論理的記憶装置を、実際にデータを記憶する物理記憶装置に配置する記憶装置において、前記論理的記憶装置の物理記憶装置への配置を動的に変更することにより記憶装置のアクセス性能を向上する技術を開示している。アクセス頻度が高い物理記憶装置に記憶されているデータの一部を前記の配置動的変更機能を用いて他の物理記憶装置に移動することにより、特定の物理記憶装置のアクセス頻度が高くならないようにし、これにより記憶装置を全体としてみたときの性能を向上させる。また、配置動的変更機能による高性能化処理の自動実行方法についても開示している。
【０００６】
特開２００１－６７１８７号公報（文献４）は、計算機がアクセスするために利用する論理的記憶装置を、実際にデータを記憶する物理記憶装置に配置し、前記論理的記憶装置の物理記憶装置への配置を動的に変更する機能を有する記憶装置において、論理的記憶装置の物理記憶装置への配置の変更案を作成する際に、物理記憶装置を属性の異なるグループに分割し、それを考慮した配置変更案を作成し、その配置変更案に従って自動的に論理的記憶装置の配置を変更する技術を開示している。配置変更案作成時に、物理記憶装置を属性毎にグループ化し、論理的記憶装置の配置先として、それが有している特徴にあった属性を保持している物理記憶装置のグループに属する物理記憶装置を選択する配置変更案を作成することによりより、良好な配置変更案を作成する。
【０００７】
【発明が解決しようとする課題】
従来の技術には以下のような問題が存在する。
【０００８】
文献１で述べられているものは、管理者がデータの配置を決定する際に考慮すべき項目である。現在、１つの記憶装置内に多数の物理記憶装置を有し、多数の計算機により共有されるものが存在する。この記憶装置においては、多くの場合、ホストが認識する論理的記憶装置を実際にデータを記憶する物理記憶装置の適当な領域に割り当てることが行われる。このような記憶装置を利用する場合、人間がすべてを把握することは困難であり、このような記憶装置を含む計算機システム側に何かしらのサポート機能が存在しなければ文献１に述べられている問題点を把握することすら困難となる。また、問題点を把握することができたとしても、計算機システム側にデータの移動機能が存在しない場合には、記憶装置上のデータの再配置のためにデータのバックアップとリストアが必要となり、その処理に多大な労力を必要とする。
【０００９】
文献２で述べられている技術では、ボリュームマネージャレイヤにおけるオブジェクトの稼動状況によるデータ再配置案を作成する機能を実現しているが、記憶装置から更に高いアクセス性能を得ようとする場合にはＤＢＭＳレイヤにおけるデータの特徴を考慮して配置を決定する必要があるがその点の解決方法については何も述べていない。
【００１０】
文献３，文献４で述べられている技術においては、記憶装置を利用するアプリケーションが利用するデータに関する特徴としてアクセス頻度とシーケンシャルアクセス性程度しか

考慮していないため、アプリケーションから見た場合に必ずしも良好な配置が実現できるわけではない。例えば、ＤＢＭＳでは表データとそれに対応する索引データを同時にアクセスすることが多いが、これらのデータを同一の物理記憶装置に配置する可能性がある。この場合、物理記憶装置においてアクセス競合が発生し、記憶装置のアクセス性能が低下する可能性がある。

【００１１】

本発明の第一の目的は、ＤＢＭＳが管理するデータの特性を考慮して記憶装置におけるデータ記憶位置を決定することにより、ＤＢＭＳに対してより好ましいアクセス性能特性を持つ記憶装置を保持する計算機システムを実現し、ＤＢＭＳの性能を向上させることである。特に、複数の記憶装置を利用するＤＢシステムにおいて、各記憶装置間へアクセス要求が適切に分散化されるようにし、ＤＢＭＳの処理性能を向上させる。

【００１２】

本発明の第二の目的は、ＤＢＭＳが稼動している計算機システムにおいて、ＤＢＭＳの特性を考慮した記憶装置の良好なアクセス性能特性達成を目的としたデータ記憶位置再配置処理を実現することにより、計算機システムの性能に関する管理コストを削減することである。

【００１３】

【課題を解決するための手段】

ＤＢＭＳに関する情報を1箇所に集約させ、そこでデータの再配置案を決定し、そのデータ再配置案に従うようにデータの移動指示を発行することにより、ＤＢＭＳに対してより好ましい性能特性を持つデータ配置を実現する。

【００１４】

計算機システム内には、複数のＤＢＭＳが稼動し、複数の記憶装置が利用されている可能性がある。そのため、ＤＢＭＳや記憶装置の構成その他の情報を一箇所に集約し、そこで集約した全情報を考慮してデータの再配置案を作成する。

【００１５】

ＤＢＭＳが管理するデータの配置決定方法として、以下のものを採用する。データ更新時に必ず書き込みが実行される更新ログを、他のデータと異なる物理記憶装置に配置して相互干渉しないようにすることによりＤＢＭＳに対して良好な性能特性を得ることができる。同時にアクセスされる可能性が極めて高い表データとそれに対応する索引データを異なる物理記憶装置に配置することによりＤＢＭＳに対して良好な性能特性を得ることができる。ＤＢＭＳに関する情報を利用して、データがシーケンシャルにアクセスされる場合のアクセス順序を予測し、その構造を保持するように物理記憶装置に記憶する。

【００１６】

現在、計算機のオペレーティングシステム（ＯＳ）・データ転送経路中のスイッチ・記憶装置内部においてデータの記憶位置を変更する技術が存在する。データの記憶位置の変更はそれらの技術を用いて実現する。そこで、前記の項目を考慮して決定されたデータ再配置案に従って、データの記憶位置の変更を管理する部分に対してデータ配置変更の指示を発行する。

【００１７】

【発明の実施の形態】

以下、本発明の実施の形態を説明する。なお、これにより本発明が限定されるものではない。

＜第一の実施の形態＞

本実施形態では、ＤＢＭＳが実行される計算機と記憶装置がスイッチを用いて接続された計算機システムにおいて、データの記憶位置の管理を行う計算機が存在し、そこで計算機システム内のデータの記憶位置の管理を行う。本実施例におけるスイッチは、記憶装置から提供される記憶領域を組み合わせて仮想的な記憶装置を作成する機能を有する。また、記憶装置においても、記憶装置内部でデータの記憶位置を動的に変更する機能を有する。

【００１８】

データ記憶位置管理を実施する計算機は、ＤＢＭＳに関する情報、データの記憶位置のマッピングに関する情報、記憶装置の構成情報を取得し、それらを用いて好適なデータ再配置案を作成する。スイッチと記憶装置に対して作成したデータ配置を指示し、それらのデータ再配置機能を用いてそのデータ再配置案に従ったデータ配置を実現する。

【００１９】

図１は、本発明の第一の実施の形態における計算機システムの構成図である。本実施の形態における計算機システムは、ＤＢホスト８０、データ位置管理サーバ８２、記憶装置１０から構成される。ＤＢホスト８０、データ位置管理サーバ８２、記憶装置１０はそれぞれが保有するネットワークインターフェイス７８を通してネットワーク７９に接続されている。また、ＤＢホスト８０、記憶装置１０はそれぞれが保有するＩ／Ｏパスインターフェイス７０からＩ／Ｏパス７１を介して仮想ボリュームスイッチ７２に接続され、これらを通して記憶装置１０とＤＢホスト８０間のデータ転送を行う。

【００２０】

本実施の形態においては、記憶装置１０とＤＢホスト８０間のデータ転送を行うＩ／Ｏパス７１とネットワーク７９を異なるものとしているが、例えばｉＳＣＳＩのような計算機と記憶装置間のデータ転送をネットワーク上で実施する技術も開発されており、本実施の形態においてもこの技術を利用してもよい。このとき、仮想ボリュームスイッチ７２において、Ｉ／Ｏパス７１とネットワーク７９との間でデータ転送が可能であれば記憶装置１０もしくはＤＢホスト８０においてＩ／Ｏパスインターフェイス７０がネットワークインターフェイス７８を兼ねても良い。

【００２１】

記憶装置１０は、記憶領域を提供するもので、その記憶領域は記憶領域管理単位であるボリュームを用いて外部に提供し、ボリューム内の部分領域に対するアクセスや管理はブロックを単位として実行する。記憶装置１０は、ネットワークインターフェイス７８、Ｉ／Ｏパスインターフェイス７０、記憶装置制御装置１２、ディスクコントローラ１６、物理記憶装置１８から構成され、ネットワークインターフェイス７８、Ｉ／Ｏパスインターフェイス７０、記憶装置制御装置１２、ディスクコントローラ１６はそれぞれ内部バス２０により接続され、ディスクコントローラ１６と物理記憶装置１８は物理記憶装置バス２２により接続される。

【００２２】

記憶装置制御装置１２は、ＣＰＵ２４とメモリ２６を有する。メモリ２６上には、記憶装置におけるキャッシュメモリとして利用するデータキャッシュ２８が割り当てられ、記憶装置を制御するためのプログラムである記憶装置制御プログラム４０が記憶される。また、メモリ２６上には、物理記憶装置１８の稼動情報である物理記憶装置稼動情報３２、記憶装置１０が提供するボリュームを物理的に記憶する物理記憶装置１８上の記憶位置の管理情報であるボリューム物理記憶位置管理情報３６を保持する。

【００２３】

図中の記憶装置１０は、複数の物理記憶装置１８を有し、１つのボリュームに属するデータを複数の物理記憶装置１８に分散配置することが可能である。また、データが記憶される物理記憶装置１８上の位置を動的に変更する機能を有する。このデータ移動指示は、ネットワークインターフェイス７８を通して外部から行うことが可能である。ただし、それらは必須のものではなく、１つの物理記憶装置の記憶領域をそのままボリュームとして提供する記憶装置１０でも本実施の形態にあてはめることができる。

【００２４】

仮想ボリュームスイッチ７２は、記憶装置１０から提供される記憶領域管理単位であるボリュームの一部領域または全領域を１つ以上組み合わせた仮想的なボリュームである仮想ボリュームをＤＢホスト８０に提供する機能を有する。仮想ボリュームスイッチ７２は、ネットワークインターフェイス７８を有し、仮想ボリュームスイッチ７２において形成されるボリュームに関する情報である仮想ボリューム情報７３を保持する。

【００２５】

仮想記憶スイッチ７２はホストから行われたアクセス要求を仮想ボリューム情報７３を参照して適切な記憶装置１０の適切な記憶領域へのアクセスへと変換してアクセス処理を実現する。また、ある仮想ボリュームのデータが記憶されるボリュームやそのボリューム内の記憶位置を動的に変更する機能を有する。このデータ移動指示は、ネットワークインターフェイス７８を通して外部から行うことが可能である。

【００２６】

ＤＢホスト８０、データ位置管理サーバ８２においては，それぞれＣＰＵ８４、ネットワークインターフェイス７８、メモリ８８を有し、メモリ８８上にオペレーティングシステム（ＯＳ）１００が記憶・実行されている。

【００２７】

ＤＢホスト８０はＩ／Ｏパスインターフェイス７０を有し、仮想ボリュームスイッチ７２により提供されるボリュームに対してアクセスを実行する。ＯＳ１００内にファイルシステム１０４と１つ以上のボリュームからホストが利用する論理的なボリュームである論理ボリュームを作成するボリュームマネージャ１０２と、ファイルシステム１０４やボリュームマネージャ１０２により、ＯＳ１００によりアプリケーションに対して提供されるファイルや論理ローボリュームに記憶されたデータの記録位置等を管理するマッピング情報１０６を有する。

【００２８】

ＯＳ１００が認識する仮想ボリュームやボリュームマネージャ１０２により提供される論理ボリュームに対して、アプリケーションがそれらのボリュームをファイルと等価なインターフェイスでアクセスするための機構であるローデバイス機構をＯＳ１００が有していても良い。図中の構成ではボリュームマネージャ１０２が存在しているが、本実施の形態においてはボリュームマネージャ１０２における論理ボリュームの構成を変更することはないので、ボリュームマネージャ１０２が存在せずにファイルシステムが直接仮想ボリュームスイッチ７２により提供される仮想ボリュームを利用する構成に対しても本実施の形態を当てはめることができる。

【００２９】

ＤＢホスト８０のそれぞれのメモリ８８上ではＤＢＭＳ１１０、データ位置管理副プログラム１２０が記憶・実行され、実行履歴情報１２２が記憶されている。ＤＭＢＳ１１０は内部にスキーマ情報１１４を有している。図中では、ＤＭＢＳ１１０が１台のホストに１つのみ動作しているが、後述するように、ＤＭＢＳ１１０毎の識別子を用いて管理を行うため、１台のホストに複数動作していても本実施の形態にあてはめることができる。

【００３０】

データ位置管理サーバ８２のメモリ８８上ではデータ位置管理主プログラム１３０が記憶・実行され、記憶装置１０内の物理記憶装置１８の稼動情報である記憶装置稼動情報１３２、各記憶装置１０における物理構成やボリュームの物理記憶位置に関する情報である記憶装置構成情報１３４、ＤＢホスト８０上のＤＢＭＳ１１０におけるスキーマに関する情報のうちデータ位置管理に必要なデータであるＤＢＭＳスキーマ情報１３６、ＤＢＭＳ１１０におけるＤＢＭＳ処理の実行履歴に関する情報であるＤＢＭＳ実行履歴情報１３８、ＤＢホスト８０におけるＯＳ１００内のマッピング情報１０６と仮想ボリュームスイッチ７２内の仮想ボリューム情報７３に関する情報を含むデータ記憶位置管理情報１４０が記憶される。

【００３１】

図中では、データ位置管理サーバ８２はＤＢホスト８０と異なる計算機となっているが、任意のＤＢホスト８０がデータ位置管理サーバ８２の役割を兼ねても本実施の形態に当てはめることができる。更に、仮想ボリュームスイッチ７２上や任意の記憶装置１０上にデータ位置管理サーバ８２が提供する機能を実装しても本実施の形態に当てはめることができる。

【００３２】

図２は記憶装置１０内に保持されている物理記憶装置稼動情報３２を示す。物理記憶装置

稼動情報３２中には、記憶装置１０が提供するボリュームの識別子であるボリューム名５０１とそのボリューム名５０１を持つボリュームのデータを保持する物理記憶装置１８の識別子である物理記憶装置名５０２、そしてボリューム名５０１を持つボリュームが物理記憶装置名５０２を持つ物理記憶装置１８に記憶しているデータをアクセスするための稼動時間のある時刻からの累積値である累積稼働時間５０３の組を保持する。

【００３３】

記憶装置１０内の記憶装置制御部１２はディスクコントローラ１６を利用して物理記憶装置１８へのデータアクセスする際の開始時刻と終了時刻を取得し、そのアクセスデータがどのボリュームに対するものかを判断して開始時刻と終了時刻の差分を稼動時間として対応するボリューム名５０１と物理記憶装置名５０２を持つデータの組の累積稼働時間５０３に加算する。この情報は、必ずしも全ての記憶装置１０で取得する必要はない。

【００３４】

図３は記憶装置１０内に保持されているボリューム物理記憶位置管理情報３６を示す。ボリューム物理記憶位置管理情報３６中には、データの論理アドレスー物理記憶装置１８における記憶位置のマッピングを管理するボリューム物理記憶位置メイン情報５１０と記憶装置１０内でのボリュームに属するデータの物理記憶位置の変更処理の管理情報であるボリュームデータ移動管理情報５１１が含まれる。

【００３５】

ボリューム物理記憶位置メイン情報５１０中には、ボリューム名５０１とそのボリューム上のデータ記憶位置であるボリューム論理ブロック番号５１２とその論理ブロックが記憶されている物理記憶装置１８の識別子である物理記憶装置名５０２と物理記憶装置１８上の記憶位置である物理ブロック番号５１４の組のデータが含まれる。ここで、ボリューム名５０１が“Ｅｍｐｔｙ”であるエントリ５１５は特殊なエントリであり、このエントリは記憶装置１０内の物理記憶装置１８の領域のうち、ボリュームに割り当てられていない領域を示し、この領域に対してデータをコピーすることによりデータの物理記憶位置の動的変更機能を実現する。

【００３６】

ボリュームデータ移動管理情報５１１はボリューム名５０１と、そのボリューム内の記憶位置を変更するデータ範囲を示す移動論理ブロック番号７８２と、そのデータが新規に記憶される物理記憶装置１８の識別子とその記憶領域を示す移動先物理記憶装置名７８３と移動先物理ブロック番号７８４、現在のデータコピー元を示すコピーポインタ７８６とデータの再コピーの必要性を管理する差分管理情報７８５の組が含まれる。

【００３７】

差分管理情報７８５とコピーポインタ７８６を用いたデータの記憶位置変更処理の概略を以下に示す。差分管理情報７８５はある一定量の領域毎にデータコピーが必要である「１」または不必要「０」を示すデータを保持する。データの記憶位置変更処理開始時に全ての差分管理情報７８５のエントリを１にセットし、コピーポインタ７８６を移動元の先頭にセットする。

【００３８】

コピーポインタ７８６にしたがって差分管理情報７８５に１がセットされている領域を順次移動先にデータをコピーし、コピーポインタ７８６を更新していく。差分管理情報７８５で管理される領域をコピーする直前に、その対応するエントリを０にセットする。データコピー中に移動領域内のデータに対する更新が行われた場合、それに対応する差分管理情報７８５のエントリを１にセットする。

【００３９】

一度全領域のコピーが完了した段階で差分管理情報７８５内のエントリが全て０になったかを確認し、全て０であればボリューム物理記憶位置メイン情報５１０を更新してデータの記憶位置変更処理は完了する。１のエントリが残っている場合には、再度それに対応する領域をコピーする処理を前記手順で繰り返す。なお、データ記憶位置の動的変更機能の実現方法は他の方式を用いても良い。この場合には、ボリューム物理記憶位置管理情報３

６中にはボリュームデータ移動管理情報５１１ではなく他のデータ記憶位置の動的変更機能のための管理情報が含まれることになる。

【００４０】

図４はＤＢホスト８０のＯＳ１００内に記憶されているマッピング情報１０６を示す。マッピング情報１０６中には、ボリュームローデバイス情報５２０、ファイル記憶位置情報５３０と論理ボリューム構成情報５４０が含まれる。ボリュームローデバイス情報５２０中にはＯＳ１００においてローデバイスを指定するための識別子であるローデバイスパス名５２１とそのローデバイスによりアクセスされる仮想ボリュームあるいは論理ボリュームの識別子であるローデバイスボリューム名５２２の組が含まれる。

【００４１】

ファイル記憶位置情報５３０中には、ＯＳ１００においてファイルを指定するための識別子であるファイルパス名５３１とそのファイル中のデータ位置を指定するブロック番号であるファイルブロック番号５３２とそれに対応するデータが記憶されている仮想ボリュームもしくは論理ボリュームの識別子であるファイル配置ボリューム名５３３とそのボリューム上のデータ記憶位置であるファイル配置ボリュームブロック番号５３４の組が含まれる。

【００４２】

論理ボリューム構成情報５４０中にはボリュームマネージャ１０２により提供される論理ボリュームの識別子である論理ボリューム名５４１とその論理ボリューム上のデータの位置を示す論理ボリューム論理ブロック番号５４２とその論理ブロックが記憶されている仮想ボリュームの識別子である仮想ボリューム名５４３と仮想ボリューム上の記憶位置である仮想ボリュームブロック番号５４４の組が含まれる。

【００４３】

図５はＤＢＭＳ１１０内に記憶されているその内部で定義・管理しているデータその他の管理情報であるスキーマ情報１１４を示す。スキーマ情報１１４には、表のデータ構造や制約条件等の定義情報を保持する表定義情報５５１、索引のデータ構造や対象である表等の定義情報を保持する索引定義情報５５２、利用するログに関する情報であるログ情報５５３、利用する一時表領域に関する情報である一時表領域情報５５４、管理しているデータのデータ記憶位置の管理情報であるデータ記憶位置情報５５５、キャッシュの構成に関する情報であるキャッシュ構成情報５５６とデータをアクセスする際の並列度に関する情報である最大アクセス並列度情報５５７を含む。

【００４４】

データ記憶位置情報５５５中には、表、索引、ログ、一時表領域等のデータ構造の識別子であるデータ構造名５６１とそのデータを記憶するファイルまたはローデバイスの識別子であるデータファイルパス名５６２とその中の記憶位置であるファイルブロック番号５６３の組が含まれる。

【００４５】

キャッシュ構成情報５５６はＤＢＭＳ１１０が三種類のキャッシュ管理のグループを定義し、そのグループに対してキャッシュを割り当てている場合を示す。キャッシュ構成情報５５６中には、グループ名５６５とグループ中のデータ構造のデータをホスト上にキャッシュする際の最大データサイズであるキャッシュサイズ５６６とそのグループに所属するデータ構造の識別子の所属データ構造名５６７の組が含まれる。最大アクセス並列度情報５５７には、データ構造名５６１とそのデータ構造にアクセスする際の一般的な場合の最大並列度に関する情報である最大アクセス並列度５６９の組が含まれる。

【００４６】

図６はＤＢホスト８０のメモリ８８上に記憶されている実行履歴情報１２２を示す。実行履歴情報１２２中には、ＤＢＭＳ１１０で実行されたクエリ５７０の履歴が記憶されている。この情報は、ＤＢＭＳ１１０が作成する。またはＤＢＭＳのフロントエンドプログラムがこの情報を作成する。この場合には、ＤＢＭＳフロントエンドプログラムが存在する計算機に実行履歴情報１２２が記憶されることになる。

【００４７】
図７は仮想ボリュームスイッチ７２が保持する仮想ボリューム情報７３を示す。仮想ボリューム情報７３は仮想ボリューム記憶位置情報７９０と仮想ボリュームデータ移動管理情報７９１を含む。仮想ボリューム記憶位置情報７９０中には、仮想ボリュームスイッチ７２がＤＢホスト８０に提供する仮想ボリュームの識別子である仮想ボリューム名５４３とその仮想ボリューム上のデータの記憶位置を示す仮想ボリュームブロック番号５４４とそのブロックが記憶されている記憶装置１０の識別子である記憶装置名５８３とそのボリュームの識別子であるボリューム名５１１とボリューム上の記憶位置であるボリューム論理ブロック番号５１２の組が含まれる。
【００４８】
仮想ボリューム名５４３が“Ｅｍｐｔｙ”であるエントリ５８５は特殊なエントリであり、このエントリに含まれる記憶装置１０上の領域はＤＢホスト８０に対して仮想ボリュームとして提供されていない領域を示す。これらの領域に対して仮想ボリュームスイッチ７２はデータの移動を行うことができる。
【００４９】
仮想ボリュームデータ移動管理情報７９１は仮想ボリューム名５４３と、そのボリューム内の記憶位置を変更するデータ範囲を示す移動仮想ボリュームブロックブロック番号７９３と、そのデータが新規に記憶される記憶装置１０の識別子とその記憶領域を示す移動先記憶装置名７９４と移動先ボリューム名７９５と移動先論理ブロック番号７９６、現在のデータコピー元を示すコピーポインタ７８６とデータの再コピーの必要性を管理する差分管理情報７８５の組が含まれる。
【００５０】
図３中のボリュームデータ移動管理情報５１１で説明したのと同様の方式によりデータの記憶位置の動的変更機能を実現できる。データ記憶位置の動的変更機能の実現方法は他の方式を用いても良い。この場合には、仮想ボリューム情報７３中には仮想ボリュームデータ移動管理情報７９１ではなく他のデータ記憶位置の動的変更機能のための管理情報が含まれることになる。
【００５１】
図８はデータ位置管理サーバ８２上に記憶される記憶装置稼動情報１３２を示す。記憶装置稼動情報１３２中には、記憶装置１０の識別子である記憶装置名５８３、記憶装置１０が提供するボリュームの識別子であるボリューム名５０１、記憶装置１０に存在する物理記憶装置１８の識別子である物理記憶装置名５０２、記憶装置名５８３とボリューム名５０１、物理記憶装置名５０２により特定される領域の前回取得時の累積稼働時間５０３の値である旧累積稼動時間５９３とある一定時間内の動作時間の割合を示す稼働率５９４の組と、稼働率計算のために前回累積稼動時間を取得した時刻である前回累積稼動時間取得時刻５９５を含む。
【００５２】
記憶装置１０は物理記憶装置稼動情報３２を外部に提供する機構を有し、それを利用して記憶位置管理主プログラム１３０は記憶装置１０で取得・記憶されている物理記憶装置稼動情報３２をネットワーク７９を通して一定間隔で取得し、取得した累積稼働時間５０３と旧累積稼働時間５９３、前回累積稼動時間取得時刻５９５と現データ取得時刻を用いて前回累積稼動時間取得時刻５９５と現データ取得時刻間の稼働率５９４を計算・記憶する。その後、取得した累積稼働時間５０３を旧累積稼働時間５９３に、現データ取得時刻を前回累積稼動時間取得時刻５９５に記憶する。
【００５３】
なお、全ての記憶装置１０で物理記憶装置稼動情報３２を保持しているとは限らない。その場合には物理記憶装置稼動情報３２を保持している記憶装置１０についてのみ記憶装置稼動情報１３２のエントリに含める。また、全ての記憶装置１０で物理記憶装置稼動情報３２を保持していない場合には記憶装置稼動情報１３２を保持しなくてもよい。
【００５４】

図９はデータ位置管理サーバ８２上に記憶される記憶装置構成情報１３４を示す。記憶装置構成情報１３４中には、記憶装置１０の識別子である記憶装置名５８３と、記憶装置１０がデータ記憶位置の動的変更機能を有しているかいないかの情報である移動機能情報６０１と記憶装置１０が保持しているデータキャッシュの容量であるデータキャッシュ容量６０２、記憶装置名５８３を持つ記憶装置１０におけるボリューム物理記憶位置メイン情報５１０を保持する記憶装置ボリューム物理記憶位置管理情報６０３の組を保持する。

【００５５】

記憶装置１０はボリューム物理記憶位置メイン情報５１０とデータキャッシュ２８のサイズに関する情報を外部に提供する機構を有し、記憶装置構成情報１３４を作成するため、データ記憶位置管理主プログラム１３０は記憶装置１０からボリューム物理記憶位置情報３６とデータキャッシュ２８のサイズに関する情報をネットワーク７９を通して取得する。記憶装置１０はデータキャッシュ２８のサイズは必ずしも外部に提供する機能を有する必要はなく、その場合には，データキャッシュ容量６０２の対応部分はデータ無効を記憶しておく。

【００５６】

一方、ボリューム物理記憶位置メイン情報５１０に関しては、記憶装置１０が外部に提供する機能を有さなくても構わない場合は、物理記憶装置１８を１つしか保持せずにそれをそのまま1つのボリュームとして提供する等、記憶装置１０が提供するボリュームがどのように物理記憶装置１８上に記憶されるかが固定され、かつ、そのマッピングをあらかじめデータ記憶位置管理主プログラム１３０が理解している場合である。このとき、データ記憶位置管理主プログラム１３０はこのルールに従って記憶装置ボリューム物理記憶位置管理情報６０３の内容を設定する。このルールは設定ファイル等を用いて管理者がデータ記憶位置管理主プログラム１３０に与える。

【００５７】

図１０はデータ位置管理サーバ８２上に記憶されるＤＢＭＳスキーマ情報１３６を示す。ＤＢＭＳスキーマ情報１３６は、ＤＢＭＳデータ構造情報６２１、ＤＢＭＳデータ記憶位置情報６２２、ＤＢＭＳパーティション化表・索引情報６２３、ＤＢＭＳ索引定義情報６２４、ＤＢＭＳキャッシュ構成情報６２５、ＤＢＭＳホスト情報６２６を含む。

【００５８】

ＤＢＭＳデータ構造情報６２１はＤＢＭＳ１１０で定義されているデータ構造に関する情報で、ＤＢＭＳ１１０の識別子であるＤＢＭＳ名６３１、ＤＢＭＳ１１０内の表・索引・ログ・一時表領域等のデータ構造の識別子であるデータ構造名５６１、データ構造の種別を表すデータ構造種別６４０、そのデータ構造をアクセスする際の最大並列度に関する情報である最大アクセス並列度５６９の組を保持する。このとき、データ構造によっては最大アクセス並列度５６９の値を持たない。

【００５９】

ＤＢＭＳデータ記憶位置情報６２２はＤＢＭＳ名６３１とそのＤＢＭＳにおけるデータ記憶位置管理情報５５５であるデータ記憶位置管理情報６３８の組を保持する。ＤＢＭＳパーティション化表・索引情報６２３は、１つの表や索引をある属性値により幾つかのグループに分割したデータ構造を管理する情報で、パーティション化されたデータ構造が所属するＤＢＭＳ１１０の識別子であるＤＢＭＳ名６３１と分割化される前のデータ構造の識別子であるパーティション元データ構造名６４３と分割後のデータ構造の識別子であるデータ構造名５６１とその分割条件を保持するパーティション化方法６４４の組を保持する。今後、パーティション化されたデータ構造に関しては、特に断らない限り単純にデータ構造と呼ぶ場合にはパーティション化後のものを指すものとする。

【００６０】

ＤＢＭＳ索引定義情報６２４には、ＤＢＭＳ名６３１、索引の識別子である索引名６３５、その索引のデータ形式を示す索引タイプ６３６、その索引がどの表のどの属性に対するものかを示す対応表情報６３７の組を保持する。ＤＢＭＳキャッシュ構成情報６２５は、ＤＢＭＳ１１０のキャッシュに関する情報であり、ＤＢＭＳ名６３１とＤＢＭＳ１１０に

おけるキャッシュ構成情報５５６の組を保持する。ＤＢＭＳホスト情報６２６は、ＤＢＭＳ名６３１を持つＤＢＭＳ１１０がどのホスト上で実行されているかを管理するもので、ＤＢＭＳ名６３１とＤＢＭＳ実行ホストの識別子であるホスト名６５１の組を保持する。

【００６１】

ＤＢＭＳスキーマ情報１３６中のＤＢＭＳホスト情報６２６以外は、データ記憶位置管理主プログラム１３０がＤＢＭＳ１１０が管理しているスキーマ情報１１４の中から必要な情報を取得して作成するものである。ＤＢＭＳ１００のスキーマ情報１１４は、ネットワーク７９を通してデータ位置管理主プログラム１３０が直接、あるいは、データ位置管理副プログラム１２０を介して、ＳＱＬ等のデータ検索言語を用いてビューとして公開されている情報を取得するか、または、専用の機構を用いて取得する。ＤＢＭＳホスト情報６２６はシステム構成情報であり、管理者が設定する。

【００６２】

図１１はデータ位置管理サーバ８２上に記憶されるＤＢＭＳ実行履歴情報１３８を示す。ＤＢＭＳ実行履歴情報１３８には、ＤＢＭＳ１１０の識別子のＤＢＭＳ名６３１と各々のＤＢＭＳ１１０で実行されたクエリ５７０の履歴が保持される。これは、データ位置管理主プログラム１３０がネットワーク７９を通してＤＢホスト８０内に記憶されている実行履歴情報１２２をデータ位置管理副プログラム１２０を利用して収集し、それを記憶したものである。

【００６３】

また、前述のように実行履歴情報１２２がＤＢＭＳフロントエンドプログラムが実行される計算機上に記憶される可能性も存在する。この場合には、ＤＢＭＳフロントエンドプログラムが実行される計算機から実行履歴情報１２２をデータ位置管理サーバ８２へ転送する手段を設け、転送された実行履歴情報１２２をデータ位置管理主プログラム１３０がＤＢＭＳ実行履歴情報１３８として記憶する。なお、本実施の形態においては、全てのＤＢＭＳ１１０から実行履歴情報１２２を収集する必要はなく、ＤＢＭＳ実行履歴情報１３８は存在しなくてもよい。

【００６４】

図１２はデータ位置管理サーバ８２上に記憶されるデータ記憶位置管理情報１４０を示す。データ記憶位置管理情報１４０には、ホストマッピング情報６５０と仮想ボリューム記憶位置管理情報７９０が含まれる。ホストマッピング情報６５０には、ホストの識別子であるホスト名６５１とそのホストにおけるマッピング情報１０６の組が保持される。これは、データ位置管理主プログラム１３０がネットワーク７９を通して、ＤＢホスト８０のＯＳ１００が保持しているマッピング情報１０６を、データ位置管理副プログラム１２０を利用して収集し、それを記憶したものである。

【００６５】

データ位置管理副プログラム１２０は、ＯＳ１００が提供している管理コマンドや情報提供機構、参照可能な管理データの直接解析等によりマッピング情報１０６を取得する。仮想ボリュームスイッチ７２は外部に仮想ボリューム記憶位置管理情報７９０を提供する機構を有し、データ位置管理主プログラム１３０はネットワーク７９を介して仮想ボリュームスイッチ７２から仮想ボリューム記憶位置管理情報７９０を取得する。

【００６６】

図１３はデータ位置管理主プログラム１３０によるデータ再配置処理の処理フローを示す。ここで、処理開始は管理者の指示によることとする。後述するように、複数の異なった種類のデータ配置解析・データ再配置案作成処理を実行可能であり、処理すべき種類の指定をして処理を開始する。また、処理にパラメータが必要な場合は併せてそれを管理者が指示をするものとする。本実施の形態においては、データの記憶位置の動的変更機能は仮想ボリュームスイッチ７２と記憶装置１０が保持する。ただし、記憶装置１０はデータの記憶位置の動的変更機能を有さなくてもよい。

【００６７】

ステップ２００１でデータ再配置処理を開始する。このとき、データ配置解析・データ再

配置案作成処理として何を実行するか指定する。また、必要であればパラメータを指定する。

【００６８】

ステップ２００２でデータ再配置処理に必要な情報を収集し、記憶装置稼動情報１３２、記憶装置構成情報１３４、ＤＢＭＳスキーマ情報１３６、ＤＢＭＳ実行履歴情報１３８、データ記憶位置管理情報１４０として記憶する。なお、このデータ収集は、ステップ２００１の処理開始とは無関係にあらかじめ実行しておくこともできる。この場合には、情報を取得した時点から現在まで情報に変更がないかどうかをこのステップで確認する。

【００６９】

ステップ２００３では、ワーク領域を確保し、その初期化を行う。ワーク領域としては、図１４に示すデータ再配置ワーク情報６７０と作成した図１５に示す移動プラン情報７５０を利用する。データ再配置ワーク情報６７０と移動プラン情報７５０の詳細とその初期データ作成方法は後述する。

【００７０】

ステップ２００４でデータ配置の解析・再配置案の作成処理を実行する。後述するように、データ配置の解析・再配置案作成処理は複数の観点による異なったものが存在し、このステップではステップ２００１で指定された処理を実行する。またステップ２００１でパラメータを受け取った場合には、それを実行する処理に与える。

【００７１】

ステップ２００５ではステップ２００４のデータ再配置案作成処理が成功したかどうか確認する。成功した場合にはステップ２００７に進む。失敗した場合にはステップ２００６に進み、管理者にデータ再配置案作成が失敗したことを通知し、ステップ２０１０に進み処理を完了する。

【００７２】

ステップ２００７では、ステップ２００４で作成されたデータ再配置案を管理者に提示する。この提示を受けて管理者はデータ再配置案に問題がないか判断する。

ステップ２００８では、データの再配置を続行するか否かを管理者から指示を受ける。続行する場合にはステップ２００９に進み、そうでない場合にはステップ２０１０に進み処理を完了する。

【００７３】

ステップ２００９では、ステップ２００４で作成されたデータの再配置案を基にデータの再配置指示を仮想ボリュームスイッチ７２、あるいは、記憶装置１０に対して発行する。仮想ボリュームスイッチ７２と記憶装置１０はネットワーク７９を通してデータのデータ再配置指示を受ける機能を有し、これを利用して指示が出される。

【００７４】

この指示の形式としては、仮想ボリュームスイッチ７２に対しては、それが提供する仮想ボリュームのあるデータ領域を、記憶装置１０のボリューム内のデータ領域を指定してそこへの移動を指示するものとなり、記憶装置１０に対しては、それが提供するボリュームのあるデータ領域を、その記憶装置１０内の物理記憶装置１８のデータ領域を指定してそこへの移動を指示するものとなる。この指示に従って、仮想ボリュームスイッチ７２、あるいは、記憶装置１０はデータの再配置処理を実行する。

【００７５】

ステップ２０１０でデータ再配置処理は完了である。

【００７６】

この処理フローは、管理者の指示で処理を開始し、作成されたデータ再配置案に問題点がないかステップ２００７と２００８で管理者が判断している。この管理者による確認を省略し、タイマを用いて設定した処理開始時刻にデータ配置解析・再配置案作成処理を開始することによりデータ再配置処理の自動化も可能である。

【００７７】

図１４は図１３のステップ２００３において作成する情報であるデータ再配置ワーク情報

６７０を示す。データ再配置ワーク情報６７０中には、仮想ボリューム物理記憶位置情報６８０とデータ構造仮想ボリューム内位置情報６９０が含まれる。

【００７８】

仮想ボリューム物理記憶位置情報６８０は仮想ボリュームスイッチ７２が提供する仮想ボリュームのデータがどの記憶装置１０内のどの物理記憶装置１８のどの位置に記憶されているかに関する情報であり、仮想ボリューム名５４３、仮想ボリュームブロック番号５４４、記憶装置名５８３、ボリューム名５０１、物理記憶装置名５０２、物理ブロック番号５１４の組で表される。仮想ボリューム理記憶位置情報６８０では、データ記憶位置管理情報１４０中の仮想ボリューム記憶位置管理情報と、記憶装置構成情報１３４中の記憶装置名５８３と記憶装置ボリューム物理記憶位置情報６０３を参照し、記憶装置１０が提供するボリュームに関して対応する部分をまとめることにより初期データが作成される。

【００７９】

仮想ボリューム名５４３が“Ｅｍｐｔｙ”であるエントリ群６８１は記憶装置１０内の物理記憶装置１８の記憶領域のうち、データ再配置のためにデータを移動することが可能な領域の集合を表し、データ再配置案作成時に、この領域の中から適切なデータの移動先を発見する。このうち、ボリューム名５０１が有効値のエントリは、仮想ボリュームスイッチ７２におけるデータ記憶位置の動的変更に利用可能な領域であり、その利用には特に制約はない。一方のボリューム名５０１が無効値のエントリは、記憶装置１０におけるデータ記憶位置の動的変更に利用可能な領域であり、その領域が属する記憶装置１０内に記憶されているデータの記憶位置変更に関してのみ利用可能である。

【００８０】

データ構造仮想ボリューム内位置情報６９０は、ＤＢＭＳ１１０が保持しているデータ構造が仮想ボリュームスイッチ７２が提供する仮想ボリュームのどこに記憶されているかを示した情報であり、ＤＢＭＳ名６３１、データ構造名５６１、データファイルパス名５６２、ファイルブロック番号５６３、仮想ボリューム名５４３、仮想ボリュームブロック番号５４４の組を保持する。この情報では、ＤＢＭＳスキーマ情報１３６中のＤＢＭＳデータ記憶位置情報６２２とＤＢＭＳホスト情報６２６とデータ記憶位置管理情報１４０中のホストマッピング情報６５０を参照し、ファイル（ローデバイス）パス、論理（仮想）ボリュームに関して対応する部分をまとめることにより初期データが作成される。

【００８１】

図１５は図１３のステップ２００４で実行されるデータ配置解析・データ再配置案作成処理により作成されるデータ移動案を格納する移動プラン情報７５０を示す。移動プラン情報７５０は、仮想ボリュームスイッチ７２に対するデータ移動指示を記憶する仮想ボリューム移動プラン情報７５１と記憶装置１０に対するデータ移動指示を記憶する物理記憶位置移動プラン情報７５２を含む。これら情報に関しては、何もデータを持たないように初期化する。

【００８２】

仮想ボリューム移動プラン情報７５１には、移動指示の実行順序を示す移動順序７６１、移動するデータを持つ仮想ボリュームとそのデータ領域示す移動仮想ボリューム名７６２と移動仮想ボリュームブロック番号７６３、そのデータの移動先の記憶装置，ボリューム，ボリューム内の記憶領域を示す移動先記憶装置名７６４と移動先ボリューム名７６５と移動先ボリューム論理ブロック番号７６６の組を保持する。

【００８３】

物理記憶位置移動プラン情報７５２には、移動指示の実行順序を示す移動順序７６１、移動するデータを持つ記憶装置１０とボリュームとそのデータ領域示す移動記憶装置名７６７と移動ボリューム名７６８と移動ボリューム論理ブロック番号７６９、そのデータの移動先の物理記憶装置とその記憶領域を示す移動先物理記憶装置名７７１と移動先物理ブロック番号７７２の組を保持する。なお、物理記憶位置移動プラン情報７５２に関しては、全ての記憶装置１０が記憶装置内の物理記憶位置の動的変更機能を有さない場合には、この情報を保持する必要はない。

【００８４】
図１３のステップ２００４で実行されるデータ配置解析・データ再配置案作成処理について説明する。前述のように、この処理には幾つかの種類が存在する。全ての処理に共通するのは逐次的にデータ再配置するためのデータ移動案を作成することである。そのため、データ移動の順番には意味があり、移動プラン情報７５０中の移動順序７６１にその順番を保持し、その順序どおりにデータ移動を行うことによりデータの再配置を実施する。また、逐次処理のため、移動後のデータ配置をもとに次のデータの移動方法を決定する必要がある。そこで、データ移動案を作成するたびにデータ再配置ワーク情報６７０をデータ移動後の配置に更新する。
【００８５】
データ再配置案作成時のデータ移動案の作成は以下のように行う。移動したいデータ量以上の連続した移動可能領域をデータ再配置ワーク情報６７０中の情報から抽出し、その中の領域を適当に選択し、設定条件や後述する制約を満たすかどうか確認をする。もし、それらを満たす場合にはそこを移動先として設定する。それらを満たさない場合には他の領域を選択し、再度それらを満たすかどうか確認をする。以下、設定条件と制約を満たす領域を検出するか、全ての移動したいデータ量以上の連続した移動可能領域が設定条件や制約を満たさないことを確認するまで処理を繰り返す。もし、全ての領域で設定条件や制約を満たさない場合にはデータ移動案の作成に失敗として終了する。
【００８６】
このときに重要なのは移動後のデータ配置において、問題となる配置を行わないことである。特にＲＤＢＭＳにおいては、特定のデータに関してはアクセスが同時に行われる可能性が高く、それらを異なる物理記憶装置１８上に配置する必要がある。
【００８７】
そこで、以下で説明する全てのデータの移動案を作成する場合には、移動するデータに含まれるデータ構造と、移動先に含まれるデータ構造を調べ、ログとその他のデータ，一時表領域とその他のデータ、表データとそれに対して作成された木構造の索引データがデータの移動後に同じ物理記憶装置１８に配置されるかどうかを確認し、配置される場合には、その配置案は利用不可能と判断する。
【００８８】
なお、あるデータ構造がどの記憶装置１０のどのボリューム上の領域に、あるいは、どの物理記憶装置１８の領域に記憶されているか、また逆に、ある物理記憶装置１８上や記憶装置１０のボリュームの領域に記憶されるデータがどのデータ構造に対応するかは、データ再配置ワーク情報６７０中の仮想ボリューム物理記憶位置情報６８０とデータ構造仮想ボリューム内位置情報６９０を仮想ボリュームに関して対応する部分を組み合わせることにより把握可能である。
【００８９】
図１６に第１のデータ配置解析・データ再配置案作成処理である、記憶装置稼動情報１３２を基にした同時アクセス実行データ構造を分離するためのデータ再配置案作成処理の処理フローを示す。本処理においては、物理記憶装置１８の稼動率が閾値を超えたものはディスクネック状態にあると判断してそれを解消するデータの移動案を作成する。
【００９０】
前述のように、記憶装置稼動情報１３２は全ての記憶装置１０中の物理記憶装置１８に関する稼動情報を必ずしも含む訳ではない。稼動情報が存在しない物理記憶装置１８に関しては、本処理におけるデータ再配置案作成の対象外とし、それらが存在しないものとして扱う。本処理は、実測値に基づいて問題点を把握し、それを解決する方法を見つけるため、より精度の高いデータ再配置案を作成するものであり、ボトルネックの自動解決方式としてデータ移動の自動化機能に組み入れても効果的に働く。
【００９１】
ステップ２１０１で処理を開始する。本処理を開始するにあたっては、どの期間の稼動率を参照するかを指定する。

【００９２】
ステップ２１０２では、物理記憶装置１８の識別子と指定期間における物理記憶装置１８の稼動率の組を記憶するワーク領域を取得し、記憶装置稼動情報１３２を参照してその情報を設定し、物理記憶装置１８の稼動率で降順にソートする。記憶装置稼動情報１３２中では、同じ物理記憶装置１８中に記憶されているデータであっても異なるボリュームのものは分離して稼動率を取得しているため、それらの総和として物理記憶装置１８の稼動率を求める必要がある。ステップ２１０３では、ステップ２１０２のソート結果をもとに物理記憶装置１８の稼動率が閾値を超えているもののリストである過負荷確認リストを作成する。このリスト中のエントリに関しても稼動率が降順になるような順序を保つようにする。

【００９３】
ステップ２１０４では、過負荷確認リスト中にエントリが存在するか確認する。エントリが存在しない場合には、もう過負荷状態の物理記憶装置１８が存在しないものとしてステップ２１０５に進みデータ再配置案作成処理成功として処理を終了する。エントリが存在する場合には、ステップ２１０６に進む。

【００９４】
ステップ２１０６では、過負荷確認リスト中の最も物理記憶装置１８の稼動率が高いものを再配置対象の物理記憶装置１８として選択する。ステップ２１０７では、再配置対象となった物理記憶装置１８内部のボリュームとその稼動率のリストを記憶装置稼動情報１３２を参照して作成し、稼動率で降順にソートする。

【００９５】
ステップ２１０８では、リスト中のあるボリュームの稼動率があらかじめ定められた閾値を超過しているかどうか確認する。全てのボリュームの稼動率が閾値を超えていない場合には、ステップ２１１３に進み、あるボリュームの稼動率がその閾値を超えている場合には、ステップ２１０９に進む。

【００９６】
ステップ２１０９においては、稼動率が閾値を超えているボリュームに関して、確認対象の物理記憶装置１８中に同時にアクセスされる可能性があるデータの組、すなわち、ログとその他のデータ，一時表領域とその他のデータ、表データとそれに対して作成された木構造の索引データがあるそのボリューム内部に記憶されているかどうか検出する処理を行う。ステップ２１１０では、ステップ２１０９における結果を確認し、同時アクセスデータ構造の組が存在する場合にはステップ２１１１に進む。同時アクセスデータ構造の組が存在しない場合には、ステップ２１１２に進む。

【００９７】
ステップ２１１１においては、同時アクセスデータ構造の組に属するデータを異なる物理記憶装置１８に記憶するためのデータ移動案を作成し、ステップ２１１４に進む。

【００９８】
ステップ２１１２においては、現在確認対象となっているのボリューム内のデータを論理ブロック番号に従って２分割し、その片方を他の物理記憶装置１８へ移動するデータ移動案を作成し、ステップ２１１４に進む。

【００９９】
ステップ２１１３においては、現在確認対象になっている物理記憶装置１８の稼動率が閾値を下回るまで、稼動率が高いボリュームから順に、その物理記憶装置１８に記憶されているボリュームを構成するデータ全体を他の物理記憶装置１８に移動するデータ移動案を作成し、ステップ２１１４に進む。

【０１００】
ステップ２１１１，２１１２，２１１３のデータ移動先を検出する際に、移動後の移動先の記憶装置の稼動率を予測する。物理記憶装置１８毎の性能差が既知の場合にはその補正を行った移動データを含む記憶装置１８上のボリュームの稼動率分、未知の場合には補正を行わない移動データを含む記憶装置１８上のボリュームの稼動率分、データ移動により

移動先の物理記憶装置１８の稼動率が上昇すると考え、加算後の値が閾値を越えないような場所へのデータの移動案を作成する。稼動率の加算分に関して、移動データ量の比率を考慮しても良いが、ここではデータ中のアクセスの偏りを考慮して移動データに全てのアクセスが集中したと考えた判断を行う。

【０１０１】

ステップ２１１４では、データの移動案の作成に成功したかどうかを確認し、失敗した場合にはステップ２１１７に進みデータの再配置案作成処理失敗として処理を終了する。成功した場合にはステップ２１１５に進む。

【０１０２】

ステップ２１１５では作成したデータ移動案を移動プラン情報７５０に追加、ステップ２１１６に進む。ステップ２１１６ではデータ再配置ワーク情報６７０を作成したデータ移動案に従って修正し、移動先記憶装置１８のステップ２１０２で作成した物理記憶装置１８毎の稼動情報情報の値を前述の移動後の稼動率判断値に修正する。その後、現在の確認対象の物理記憶装置１８を過負荷確認リストから削除し、ステップ２１０４に戻り次の確認を行う。

【０１０３】

次に第２のデータ配置解析・データ再配置案作成処理である、ＤＢＭＳ実行履歴情報１３８を基にした同時アクセス実行データ構造を分離するためのデータ再配置案作成処理を示す。本処理においては、クエリの実行履歴から同時にアクセスされるデータの組を取得し、それらを異なる物理記憶装置１８に配置するデータ再配置案を作成する。前述のように、全てのＤＢＭＳ１１０について実行履歴を取得できるわけではない。本処理実行時に実行履歴が存在しないＤＢＭＳ１１０が利用するデータに関してはデータ再配置の対象外とする。

【０１０４】

本処理においては、図１７に示すクエリ実行時同時アクセスデータカウント情報７００を利用する。クエリ実行時同時アクセスデータカウント情報７００は、ＤＢＭＳ名６３１、同時にアクセスされる可能性のあるデータ構造のデータ構造名５６１の組を示すデータ構造名Ａ７０１とデータ構造名Ｂ７０２、そして、ＤＢＭＳ実行履歴情報１３８の解析によりそのデータ構造の組がアクセスされたと判断された回数であるカウント値７０３の組で表される。この組はカウント値７０３の値でソートしておく。

【０１０５】

クエリ実行時同時アクセスデータカウント情報７００はＤＢＭＳ実行履歴情報１３８から作成する。最初にクエリ実行時同時アクセスデータカウント情報７００のエントリを全消去する。ＤＢＭＳ１００において定型処理が行われる場合には、まず、その型により分類し、その型の処理が何回実行されたかを確認する。

【０１０６】

続いてＤＢＭＳ１００から型毎のクエリ実行プランを取得する。そのクエリ実行プランにより示される処理手順から同時にアクセスされるデータ構造の組を判別する。そして、クエリ実行時同時アクセスデータカウント情報７００中のＤＢＭＳ名６３１・データ構造名Ａ７０１・データ構造名Ｂ７０２を参照し、既に対応するデータ構造の組が存在している場合には先に求めたその型の処理回数をカウント値７０３に加算する。既に対応するデータ構造の組が存在していない場合には、新たにエントリを追加してカウント値７０３を先に求めたその型の処理回数にセットする。

【０１０７】

ＤＢＭＳ１００において非定型処理が行われる場合には、１つ１つの実行されたクエリに関してクエリ実行プランを取得し、そのクエリ実行プランにより示される処理手順から同時にアクセスされるデータ構造の組を判別する。そして、クエリ実行時同時アクセスデータカウント情報７００中のＤＢＭＳ名６３１・データ構造名Ａ７０１・データ構造名Ｂ７０２を参照し、既に対応するデータ構造の組が存在している場合にはカウント値７０３に１を加算する。既に対応するデータ構造の組が存在していない場合には、新たにエントリ

を追加してカウント値７０３に１をセットする。

【０１０８】

クエリ実行プランから同時にアクセスされる可能性があるデータ構造の判別は以下のように行う。まず、木構造の索引に対するアクセスが実施される場合には、その木構造索引データと、その索引が対象とする表データが同時にアクセスされると判断する。また、データの更新処理や挿入処理が行われる場合には、ログとその他のデータが同時にアクセスされると判断する。以下はＤＢＭＳ１１０の特性に依存するが、例えば、クエリ実行プラン作成時にネストループジョイン処理を多段に渡り実行する計画を作成し、それらの多段に渡る処理を同時に実行するＲＤＢＭＳが存在する。このＲＤＢＭＳを利用する場合にはその多段に渡るネストループジョイン処理で利用する表データとその表に対する木構造の索引データは同時にアクセスされると判断できる。

【０１０９】

このように、クエリ実行計画による同時アクセスデータの判断に関しては、ＤＢＭＳ１１０の処理特性を把握して判断する必要があるが、ここでは、対象とするＤＢＭＳ１１０の種類を絞りデータ位置管理主プログラム１３０がＤＢＭＳ１１０特有の同時アクセスデータ構造の組を把握できる機能を有することを仮定する。

【０１１０】

図１８にＤＢＭＳ実行履歴情報１３８を基にした同時アクセス実行データ構造を分離するためのデータ再配置案作成処理処理フローを示す。ステップ２２０１で処理を開始する。ステップ２２０２で実行履歴から同時にアクセスされるデータ構造の組とその実行頻度に関する情報である、前述のクエリ実行同時アクセスデータ構造カウント情報７００を作成する。ステップ２２０３において、カウント値７０３の全エントリの総和に対してカウント値７０３の値が一定割合以上のデータ構造とその所属するＤＢＭＳ１１０の組を求め、それらを確認リストとして記憶する。

【０１１１】

ステップ２２０４でステップ２２０３で求めた確認リスト中に含まれるデータ構造の組に関して、それらを異なる物理記憶装置１８に記憶するデータ再配置案を作成し、ステップ２２０５に進む。なお、ステップ２２０４の処理に関しては、図１９を用いて後で説明する。ステップ２２０５では、ステップ２２０４においてデータ再配置案の作成に成功したかどうかを確認し、成功した場合にはステップ２２０６に進みデータ再配置案作成処理成功として処理を終了し、失敗した場合にはステップ２２０７に進みデータ再配置案作成処理失敗として処理を終了する。

【０１１２】

図１９に、指定されたデータ構造とそのデータ構造と同時にアクセスされる可能性が高いデータ構造の組を分離するデータ再配置案を作成する処理のフローを示す。本処理を開始するときには、データ構造名と物理記憶装置１８から分離するデータ構造名の組のリストである確認リストを与える。

【０１１３】

ステップ２３０１で処理を開始する。ステップ２３０３で確認リスト中にエントリが存在するか確認し、存在しない場合にはステップ２３０４に進みデータ再配置案作成処理成功として処理を終了する。存在する場合にはステップ２３０５に進む。

【０１１４】

ステップ２３０５においては、確認リストから１つ確認対象データ構造名とその所属ＤＢＭＳ名の組とその分離データ構造名とその所属ＤＢＭＳ名の組の組を取得し、ステップ２３０６に進む。

【０１１５】

ステップ２３０６においては、確認対象データ構造とその分離するデータ構造が同一の物理記憶装置上に記憶されているかどうかの確認を行う。前述のように、この確認はデータ再配置ワーク情報６７０を参照することにより可能である。両データ構造が全て異なる物理記憶装置上に存在する場合にはステップ２３１２に進み、ある物理記憶装置上に両デー

タ構造が存在する場合にはステップ２３０７に進む。

【０１１６】

ステップ２３０７においては、同一の物理記憶装置上に両データ構造が存在する部分に関してそれを分離するデータ移動案を作成する。ステップ２３０８においては、そのデータ移動案作成が成功したかどうか確認し、成功した場合にはステップ２３１０に進み、失敗した場合にはステップ２３０９に進みデータ再配置案作成処理失敗として処理を終了する。

【０１１７】

ステップ２３１０においては、作成されたデータ移動案を移動プラン情報７５０に記憶する。ステップ２３１１においては、作成されたデータ移動案に従ってデータ再配置ワーク情報６７０を更新し、ステップ２３１２に進む。

【０１１８】

ステップ２３１２においては、確認リストから現在確認対象となっているデータ構造の組のエントリを削除し、ステップ２３０３に進む。

【０１１９】

図２０に、第３のデータ配置解析・データ再配置案作成処理である、データ構造の定義を基にした同時アクセス実行データ構造を分離するためのデータ再配置案作成処理の処理フローを示す。本処理においては、同時にアクセスされる可能性が高い、ログとその他のデータ、一時表領域とその他のデータ、表データとそれに対して作成された木構造の索引データが同一物理記憶装置１８上に記憶されている部分が存在しないか確認をし，そのような部分が存在する場合にはそれを解決するデータ再配置案を作成する。

【０１２０】

ステップ２４０１で処理を開始する。ステップ２４０２では、ＤＢＭＳデータ構造情報６２１を参照して全てのログであるデータ構造名５６１とそれを利用するＤＢＭＳ１１０のＤＢＭＳ名６３１の組を取得する。そして、そのデータ構造名とログ以外のデータを分離するデータ構造とする確認リストを作成し、ステップ２４０３に進む。

【０１２１】

ステップ２４０３ではステップ２４０２で作成した確認リストを用いてステップ２３０１から開始されるデータ構造分離のためのデータ再配置案作成処理を実行する。ステップ２４０４ではステップ２４０３におけるデータ再配置案作成処理が成功したか確認をし、成功した場合にはステップ２４０５に進む。失敗した場合にはステップ２４１２に進みデータ再配置案作成処理失敗として処理を終了する。

【０１２２】

ステップ２４０５では、ＤＢＭＳデータ構造情報６２１を参照して全ての一時表領域であるデータ構造名５６１とそれを利用するＤＢＭＳ１１０のＤＢＭＳ名６３１の組を取得する。そして、そのデータ構造名と一時表領域以外のデータを分離するデータ構造とする確認リストを作成し、ステップ２４０６に進む。

【０１２３】

ステップ２４０６ではステップ２４０５で作成した確認リストを用いてステップ２３０１から開始されるデータ構造分離のためのデータ再配置案作成処理を実行する。ステップ２４０７ではステップ２４０６におけるデータ再配置案作成処理が成功したか確認をし、成功した場合にはステップ２４０８に進む。失敗した場合にはステップ２４１２に進みデータ再配置案作成処理失敗として処理を終了する。

【０１２４】

ステップ２４０８では、ＤＢＭＳ索引定義情報６２４を参照して全ての木構造索引の索引名６３５とそれに対応する表のデータ構造名を対応表情報６３７から取得する。そして、索引名６３５と対応する表のデータ構造名とそれらを保持するＤＢＭＳ１１０のＤＢＭＳ名６３１を組とする確認リストを作成し、ステップ２４０９に進む。

【０１２５】

ステップ２４０９ではステップ２４０８で作成した確認リストを用いてステップ２３０１

から開始されるデータ構造分離のためのデータ再配置案作成処理を実行する。ステップ２４１０ではステップ２４０９におけるデータ再配置案作成処理が成功したか確認をし、成功した場合にはステップ２４１１に進み、データ再配置案作成処理成功として処理を終了する。失敗した場合にはステップ２４１２に進みデータ再配置案作成処理失敗として処理を終了する。

【０１２６】

図２１に第４のデータ配置解析・データ再配置案作成処理である、特定の表や索引の同一データ構造に対するアクセス並列度を考慮したデータ再配置案作成処理の処理フローを示す。この処理は、ランダムアクセス実行時の処理の並列度を考慮してディスクネックの軽減を図るためにデータの再配置を行うものである。この処理を実行する際には、データ再配置の確認対象とするデータ構造をＤＢＭＳ名６３１とデータ構造名５６１の組として指定する。

【０１２７】

ステップ２５０１で処理を開始する。ステップ２５０２において、指定されたデータ構造の物理記憶装置上に割り当てられた記憶領域利用総量を求める。この値は、指定データ構造がデータファイル上に割り当てられた容量と等しいため、ＤＢＭＳデータ記憶位置情報６２２中のデータ記憶位置を参照してその容量を求めればよい。

【０１２８】

ステップ２５０３においては、ＤＢＭＳデータ構造情報６２１を参照して指定データ構造における最大アクセス並列度５６９を取得する。ステップ２５０４において、ステップ２５０２で求めた指定データ構造の記憶領域利用総量をステップ２５０３で求めた最大アクセス並列度５６９で割った値を、指定データ構造の１つの物理記憶装置１８上への割り当てを許可する最大量として求める。この制約により、特定の物理記憶装置１８に偏ることなく最大アクセス並列度５６９以上の台数の物理記憶装置１８に指定データ構造が分散して記憶されることになり、最大アクセス並列度５６９による並列度でランダムアクセスが実行されてもディスクネックになりにくい状況となる。なお、割り当て許可最大量の値は、実際のアクセス特性を考慮してこの方法で求めた値から更に増減させても構わない。

【０１２９】

ステップ２５０５において、指定データ構造のデータがステップ２５０４で求めた最大量を超えて１つの物理記憶装置１８上に割り当てられているものが存在するかデータ再配置ワーク情報６７０を用いて確認し、そのようなものが存在しない場合にはステップ２５０９に進み、データ再配置案作成処理成功として処理を終了する。存在する場合にはステップ２５０６に進む。

【０１３０】

ステップ２５０６においては、ステップ２５０４で求めた最大量を超えて１つの物理記憶装置１８上に割り当てられている部分を解消するデータ移動案を作成する。このとき、移動案作成に考慮するデータ移動量は指定データ構造の現在の物理記憶装置１８上への割り当て量のステップ２５０４で求めた最大量からの超過分以上である必要がある。また、移動先物理記憶装置１８においても、移動後にステップ２５０４で求めた最大量を超過しないようにする必要がある。

【０１３１】

ステップ２５０７においては、ステップ２５０６のデータ移動案作成処理が成功したか確認をする。成功した場合にはステップ２５０８に進む。失敗した場合にはステップ２５１０に進み、データ再配置案作成処理失敗として処理を終了する。

【０１３２】

ステップ２５０８においては作成したデータ移動案を移動プラン情報７５０に記憶し、ステップ２５０９に進みデータ再配置案作成処理成功として処理を終了する。

【０１３３】

図２２に第５のデータ配置解析・データ再配置案作成処理である、特定の表データに対するシーケンシャルアクセス時のディスクネックを解消するデータ再配置案作成処理の処理

フローを示す。この処理を実行する際には、データ再配置の確認対象とする表をＤＢＭＳ名６３１とデータ構造名５６１の組として指定する。

【０１３４】

ＤＢＭＳ１１０毎により、シーケンシャルアクセス方法が定まっている。そこで、対象とするＤＢＭＳ１１０の種類を絞り、あらかじめデータ位置管理主プログラム１３０がＤＢＭＳ１１０におけるシーケンシャルアクセス方法を把握し、それらに対する最適化を行う。本実施の形態のＤＢＭＳ１１０におけるシーケンシャルアクセス方法は以下の方法に従うものとする。あるデータ構造のデータをシーケンシャルアクセスする場合に、データ構造が記憶されているデータファイル名５６２とファイルブロック番号５６３を昇順にソートしその順序でアクセスを実行する。

【０１３５】

その他にシーケンシャルアクセス方法の決定方法としては、データファイルを管理する内部通番とファイルブロック番号５６３の組を昇順にソートした順番にアクセスする方法等が存在し、それらを利用したシーケンシャルアクセス方法の判断を実施してもよい。

【０１３６】

また、並列にシーケンシャルアクセスを実行する場合に、その領域の分割法は並列にアクセスしない場合のシーケンシャルにアクセスする順番を並列度に合わせて等分に分割するものとする。

【０１３７】

この並列アクセスによる分割後の１つのアクセス領域を全て同一の物理記憶装置１８上に配置するのは必ずしも現実的ではない。そこで、分割後のアクセス領域がある一定量以上連続にまとまって１つの物理記憶装置上に記憶されていればよいと判断する。ただし、どのような場合でも連続してアクセスされることがなく、分割後のアクセス領域が異なるものに分類されるものに関しては、並列シーケンシャルアクセス時にアクセスがぶつかる可能性があるため、異なる物理記憶装置１８に記憶するという指針を設けて、これに沿うようなデータ配置を作成することによりシーケンシャルアクセスの性能を高める。

【０１３８】

ステップ２６０１で処理を開始する。ステップ２６０２において、指定された表の物理記憶装置上に割り当てられた記憶領域利用総量を求める。この値は、指定データ構造がデータファイル上に割り当てられた容量と等しいため、ＤＢＭＳデータ記憶位置情報６２２中のデータ記憶位置を参照してその容量を求めればよい。ステップ２６０３においては、ＤＢＭＳデータ構造情報６２１を参照して指定データ構造における最大アクセス並列度５６９を取得する。

【０１３９】

ステップ２６０４において、ステップ２６０２で求めた指定表の記憶領域利用総量をステップ２６０３で求めた最大アクセス並列度５６９で割った量が、並列アクセス時にシーケンシャルにアクセスされる１つの領域のデータ量である。そこで、データ位置管理主プログラム１３０が把握している前述のシーケンシャルアクセス先の決定方法に基づいてＤＢＭＳデータ記憶位置情報６２２中の指定データ構造のデータファイルにおける記憶先を求め、それらのアクセス方法を前述のように予測し、その結果をもとに最大アクセス並列度５６９の並列アクセスが実行されると仮定した前述のデータ分割指針を作成する。

【０１４０】

ステップ２６０５において、データ再配置ワーク情報６７０を参照しながら、指定データ構造はステップ２６０４で作成したデータ分割指針に沿ったデータ配置が物理記憶装置１８上で行われているか確認し、そうであればステップ２６０９に進み、データ再配置案作成処理成功として処理を終了する。そうでない場合にはステップ２６０６に進む。

【０１４１】

ステップ２６０６においては、物理記憶装置１８上において、ステップ２６０４で求めたデータ分割指針に従うデータ配置を求める。このとき、データがある一定値以下の領域に細分化されている場合には、連続した空き領域を探し、そこにアクセス構造を保つように

データを移動するデータ移動案を作成する。また、最大アクセス並列度５６９の並列アクセスにより異なるアクセス領域に分離されるデータが同じ物理記憶装置１８上に配置されないようなデータ移動案を作成する。

【０１４２】

ステップ２６０７においては、ステップ２６０６のデータ移動案作成処理が成功したか確認をする。成功した場合にはステップ２６０８に進み、失敗した場合にはステップ２６１０に進み、データ再配置案作成処理失敗として処理を終了する。

【０１４３】

ステップ２６０８においては、作成したデータ移動案を移動プラン情報７５０に記憶し、ステップ２６０９に進みデータ再配置案作成処理成功として処理を終了する。

【０１４４】

次に第６のデータ配置解析・データ再配置案作成処理である、特定のデータ構造に対する記憶装置１０におけるキャッシュ効果を考慮したデータ再配置案作成処理の説明を行う。この処理を実行する際には、データ再配置の確認対象とするデータ構造としてＤＢＭＳ名６３１とデータ構造名５６１を指定する。また、処理実行時に、記憶装置におけるキャッシュ効果が存在するかどうかを明示的に指定してもよい。前述のように、記憶装置構成情報１３４において、すべての記憶装置１０がデータキャッシュ容量６０２に有効な値を保持しているわけではない。そのため、データキャッシュ容量６０２が無効値である記憶装置１０は本処理の対象外とする。

【０１４５】

本処理においては、指定データ構造には記憶装置１０のキャッシュ効果が存在するかどうかを判断する。まず、データ構造の単位データ量あたりのホストキャッシュにおける平均的キャッシュ利用量を計算する。その結果を用いて十分な量のホストキャッシュが利用可能かを判断する。十分な量のホストキャッシュを利用可能な場合には、アクセス頻度が低いデータに対してのみ記憶装置からデータを読み出すことになり、記憶装置におけるキャッシュ効果は極めて低いものとなる。

【０１４６】

その確認処理のフローを図２３に示す。ステップ２８０１で処理を開始する。ステップ２８０２でＤＢＭＳキャッシュ構成情報６２５を参照して指定データ構造が属するキャッシュグループを求め、それからＤＢＭＳデータ記憶位置情報を参照してそのキャッシュグループに属するデータ構造の記憶のために割り当てられた領域量の総和を求める。

【０１４７】

ステップ２８０３において、ＤＢＭＳキャッシュ構成情報６２５から指定データ構造が属するキャッシュグループに割り当てられたキャッシュサイズ５６６を求め、それとステップ２８０２で求めた領域量の総和から指定データ構造のホストにおける単位データ量あたりの平均キャッシュ利用量を求め、それをキャッシュ効果判断閾値と比較する。単位データ量あたりの平均キャッシュ利用量が閾値以上の場合にはステップ２８０４に進み、指定データ構造は記憶装置におけるキャッシュ効果がないと判断し、ステップ２８０６に進み処理を終了する。閾値未満の場合にはステップ２８０５に進み、指定データ構造は記憶装置におけるキャッシュ効果があると判断し、ステップ２８０６に進み処理を終了する。

【０１４８】

本処理において、記憶装置１０におけるデータ構造のキャッシュ利用量を判断する。このとき、実際のデータのキャッシュ利用量はアクセスパターンに依存するが、ここでは平均的なケースを考え、データ構造のキャッシュ利用量は、その記憶装置１０に割り当てられたデータ量に比例するとする。つまり、記憶装置構成情報１３４から記憶装置ボリューム物理記憶位置管理情報６０３を参照して記憶装置１０における全データ記憶容量を求め、また、データキャッシュ容量６０２からデータキャッシュ容量を求める。これらの値から単位データ容量あたりのデータキャッシュ量が求まりこの値を元に判断する。また、記憶装置１０において、提供するボリュームをいくつかにグループ化し、それぞれで単位容量あたりのキャッシュ利用量を変化させる制御を行うことが可能である。この場合には、ボ

リューム等の記憶領域毎にデータ構造のキャッシュ利用量が変化することになる。

【０１４９】

図２４に特定のデータ構造に対する記憶装置１０におけるキャッシュ効果を考慮したデータ再配置案作成処理の処理フローを示す。ステップ２７０１で処理を開始する。ステップ２７０２において、データ再配置案の作成対象として指定されたデータ構造に対して、処理開始時にキャッシュ効果があるかないかが明示的に指定されたか確認する。明示的に指定されていない場合にはステップ２７０３に進み、明示的に指定された場合にはステップ２７０４に進む。

【０１５０】

ステップ２７０３において、前述した指定データ構造の記憶装置１０におけるキャッシュ効果が存在するかどうか確認処理を実行する。

【０１５１】

ステップ２７０４において、指定データ構造は記憶装置１０においてキャッシュ効果があるかないかを確認する。キャッシュ効果があると判断された場合にはステップ２７０５に進み、ないと判断された場合にはステップ２７０６へ進む。

【０１５２】

ステップ２７０５においては、指定データ構造のデータを十分な量のキャッシュを利用可能な装置、ボリューム等の記憶領域に移動するデータ移動案を作成する。ここで、十分な量のキャッシュを利用可能な記憶領域とは、原則として単位容量あたりのキャッシュ利用量の大きな記憶領域を意味するが、以下の意味も併せ持つ。指定データ構造のホストキャッシュの利用可能量が大きなものである場合、アクセス頻度が高いものはホストキャッシュに留まることになる。そのため、記憶装置１０におけるキャッシュ利用可能容量は、ホストキャッシュにおける利用可能量に比べて、ある一定比率以上の容量がないとその効果は小さなものと考えられる。そこで、データ構造の単位データ量あたりのホストキャッシュにおける平均的キャッシュ利用量と記憶領域における単位容量あたりの平均的キャッシュ利用量の割合を計算し、その値と判断閾値を比較して記憶領域の方がよりキャッシュを多く利用できると判断される場合でないと十分な量のキャッシュが利用可能であると判断しない。本処理終了後、ステップ２７０７に進む。

【０１５３】

ステップ２７０６においては、指定データ構造のデータをキャッシュ利用量を少量に抑えられる領域にデータを移動する、つまり、記憶装置おける単位容量あたりの平均的キャッシュ利用量が小さな装置、ボリューム等の領域へデータを移動するデータの移動案を作成し、ステップ２７０７に進む。

【０１５４】

ステップ２７０７においては、データ移動案の作成に成功したかどうかを確認する。成功した場合にはステップ２７０８に進み、失敗した場合にはステップ２７１０に進み、データ再配置案作成処理失敗として処理を終了する。

【０１５５】

ステップ２７０８においては作成したデータ移動案を移動プラン情報７５０に記憶し、ステップ２７０９に進みデータ再配置案作成処理成功として処理を終了する。

＜第二の実施の形態＞

本実施形態では、ＤＢＭＳが実行される計算機と記憶装置が接続された計算機システムにおいて、データの記憶位置の管理を行う計算機が存在し、そこで計算機システム内のデータの記憶位置の管理を行う。計算機上で実行されるＯＳ中のファイルシステムやボリュームマネージャは、動的にデータの記憶位置を変更する機能を有する。また、記憶装置においても、記憶装置内部でデータの記憶位置を動的に変更する機能を有する。

【０１５６】

データ記憶位置管理を実施する計算機は、ＤＢＭＳに関する情報、データの記憶位置のマッピングに関する情報、記憶装置の構成情報を取得し、それらを用いて好適なデータ再配置案を作成する。ファイルシステム，ボリュームマネージャ、記憶装置に対して作成した

データ配置を指示し、それらのデータ再配置機能を用いてそのデータ再配置案に従ったデータ配置を実現する。

【０１５７】

図２５は、本発明の第二の実施の形態における計算機システムの構成図である。図示されたように、本発明の第二の実施の形態は本発明の第一の実施の形態と以下の点が異なる。

【０１５８】

仮想ボリュームスイッチ７２がＩ／Ｏパススイッチ７２ｂに変更され、ネットワーク７９とは未接続になる。ＤＢホスト８０ｂで実行されるＯＳ１００が有するボリュームマネージャ１０２がボリュームマネージャ１０２ｂに、ファイルシステム１０４がファイルシステム１０４ｂに変更され、ＯＳ１００が保持するマッピング情報１０６がマッピング情報１０６ｂへとその内容が若干変更される。データ位置管理サーバ８２内に記憶されるデータ記憶位置管理情報１４０がデータ記憶位置管理情報１４０内のホストマッピング情報６５０の内容を若干変更したホストマッピング情報６５０ｂに変更される。

【０１５９】

Ｉ／Ｏパススイッチ７２ｂは仮想ボリュームスイッチ７２と比較して経路制御の機能のみを保持する。本実施の形態においては、記憶装置１０とＤＢホスト８０ｂ間のデータ転送を行うＩ／Ｏパス７１とネットワーク７９を異なるものとしているが、例えばｉＳＣＳＩのような計算機と記憶装置間のデータ転送をネットワーク上で実施する技術も開発されおり、本実施の形態においてもこの技術を利用してもよい。このとき、記憶装置１０とＤＢホスト８０ｂにおいてＩ／Ｏパスインターフェイス７０が省かれ、計算機システム内からＩ／Ｏパス７１とＩ／Ｏパススイッチ７２ｂが省かれる構成となる。

【０１６０】

ボリュームマネージャ１０２ｂはボリュームマネージャ１０２と比べて、ボリュームマネージャ１０２ｂが提供する論理ボリュームの指定領域に記憶されているデータを、指定した記憶装置１０が提供するボリュームの指定した記憶領域に移動する機能を有する。直接的にこの機能を有していない場合でも、管理領域の動的変更機能と管理領域内管理単位毎のデータ移動機能を有するものに関してはそれらの機能の組み合わせで実現できる。このデータ移動機能に関しては、ホスト上で実行される管理コマンドにより処理が実施される。

【０１６１】

ファイルシステム１０４ｂは、ファイルシステム１０４と比べて、ファイルシステム１０４ｂが管理する領域の中でデータが記憶されていない領域に対してファイルデータの一部をそこに移動する機能を有する。そのデータ移動指示方法は、移動するファイルとそのデータ領域と、移動先領域を指定するものとする。このデータ移動機能に関しては、ホスト上で実行される管理コマンドにより処理が実施される。

【０１６２】

ファイルのデータ記憶位置の動的変更機能は例えばファイルシステムのデフラグ機能として実現されており、その技術を用いてデータの移動先領域を指定できるようにすることにより前記データ移動機能を実現できる。データの移動先となることができる領域は、マッピング情報１０６ｂを参照することにより把握することができる。

【０１６３】

なお、本実施の形態においては、ボリュームマネージャ１０２ｂとファイルシステム１０４ｂのどちらか一方がデータ移動機能を有していれば良い。また、ファイルシステム１０４ｂがデータ移動機能を有している場合には、ボリュームマネージャ１０２ｂが存在しなくても本実施の形態に当てはめることができる。

【０１６４】

図２６はＯＳ１００内に記憶されるマッピング情報１０６ｂを示す。図４のマッピング情報１０６からの変更点の概略は、マッピング情報１０６においては仮想ボリュームスイッチ７２が提供する仮想ボリュームを利用していた部分をマッピング情報１０６ｂでは記憶装置１０が提供するボリュームに変更された部分と、ファイル記憶位置情報５３０では保

持していなかったファイルシステムＩＤ５３５と空き領域の管理情報がファイル記憶位置情報５３０ｂに追加されたことである。

【０１６５】

マッピング情報１０６ｂ中には、ボリュームローデバイス情報５２０ｂ、ファイル記憶位置情報５３０ｂと論理ボリューム構成情報５４０ｂが含まれる。ボリュームローデバイス情報５２０ｂ中にはＯＳ１００においてローデバイスを指定するための識別子であるローデバイスパス名５２１と、ローデバイスとして記憶装置１０が提供するボリュームを利用する場合の記憶装置１０の識別子である記憶装置名５８３、そのローデバイスによりアクセスされるボリュームあるいは論理ボリュームの識別子であるローデバイスボリューム名５２２ｂの組が含まれる。

【０１６６】

ファイル記憶位置情報５４０ｂ中には、ＯＳ１００においてファイルを指定するための識別子であるファイルパス名５３１とそのファイルが存在するファイルシステムのホスト内識別子であるファイルシステムＩＤ５３５とそのファイル中のデータ位置を指定するブロック番号であるファイルブロック番号５３２とそれに対応するデータが記憶されている記憶装置１０が提供するボリュームもしくは論理ボリュームの識別子であるファイル配置ボリューム名５３３ｂと記憶装置１０が提供するボリュームを利用する場合の記憶装置名５８３とそのボリューム上のデータ記憶位置であるファイル配置ボリュームブロック番号５３４の組が含まれる。ファイルパス名５３１が“Ｅｍｐｔｙ”であるエントリ５３６は特殊なエントリであり、これはファイルシステム内における空き領域を示すデータである。この領域に対してデータの移動を行うことができる。

【０１６７】

論理ボリューム構成情報５４０ｂ中にはボリュームマネージャ１０２ｂにより提供される論理ボリュームの識別子である論理ボリューム名５４１とその論理ボリューム上のデータの位置を示す論理ボリュームブロック番号５４２とその論理ブロックが記憶されている記憶装置１０が提供するボリュームの識別子であるボリューム名５０１とそれを提供する記憶装置名５８３とボリューム上の記憶位置であるボリューム論理ブロック番号５１２の組が含まれる。

【０１６８】

図２７はデータ位置管理サーバ８２上に記憶されるホストマッピング情報６５０ｂを示す。図１２のホストマッピング情報６５０からの変更点は、各ホストのマッピング情報１０６を保持するマッピング情報６５２が各ホストのマッピング情報１０６ｂを保持するマッピング情報６５２ｂになったことである。

【０１６９】

データ再配置処理において、データ移動機能を保持する部分が仮想ボリュームスイッチ７２からボリュームマネージャ１０２ｂ、ファイルシステム１０４ｂに変更されたことによる変更点は以下のようなものである。

【０１７０】

ボリュームマネージャ１０２ｂ、ファイルシステム１０４ｂにおいては、ＤＢホスト８０ｂ上において管理コマンドを実施することによりデータの移動を実施する。そこで、データ位置管理主プログラム１３０がネットワーク７９を通してＤＢホスト８０ｂ上で実行されているデータ位置管理副プログラム１２０にボリュームマネージャ１０２ｂ、ファイルシステム１０４ｂにおいてデータの移動を実施する管理コマンドを実行する指示を出し、それに従ってデータ位置管理副プログラム１２０が管理コマンドを実施することによりデータの移動を実施する。

【０１７１】

なお、ボリュームマネージャ１０２ｂにおいてデータ移動先として利用する領域は、データ位置管理主プログラム１３０移動指示を発行したときにはボリュームマネージャ１０２ｂの管理下に存在しない可能性がある。この場合には、データ位置管理副プログラム１２０は、データ移動の管理コマンドを実行する前に移動先領域をボリュームマネージャ１０

２ｂの管理下に収めるための管理コマンドを実行する。また、データ移動処理が終了後、他のＤＢホスト８０上のボリュームマネージャ１０２ｂがデータ移動元の領域を新たなデータ移動先として利用するため、データ移動元領域の開放のための管理コマンドを実施する。

【０１７２】

また、データ再配置案作成処理においては、データ再配置案を作成する際に利用するワーク領域であるデータ再配置ワーク情報６７０がデータ再配置ワーク情報６７０ｂに、移動プラン情報７５０が移動プラン情報７５０ｂに変更される。

【０１７３】

図２８にデータ再配置ワーク情報６７０ｂを示す。データ再配置ワーク情報６７０ｂ中には、ワーク用記憶装置ボリューム記憶位置情報６８２とワーク用空き領域情報６８３とデータ構造仮想ボリューム内位置情報６９０ｂが含まれる。図１４のデータ再配置ワーク情報６７０と比べて、仮想ボリューム物理記憶位置情報６８０はワーク用記憶装置ボリューム記憶位置情報６８２とワーク用空き領域情報６８３に分離・変更され、データ構造仮想ボリューム内位置情報６９０はデータ構造仮想ボリューム内位置情報６９０ｂに変更される。

【０１７４】

ワーク用記憶装置ボリューム記憶位置情報６８２は記憶装置１０が提供するボリュームとその物理記憶装置１８における記憶位置の一覧であり、記憶装置名５８３、ボリューム名５０１、ボリューム論理ブロック番号５１２、物理記憶装置名５０２、物理ブロック番号５１４の組を保持する。このデータは、記憶装置構成情報１３４を参照して初期化する。

【０１７５】

ワーク用空き領域情報６８３は、データ再配置案を作成する際にデータの移動先となりうる場所を管理するものであり、ホスト名６３１、ファイルシステムＩＤ５３５、論理ボリューム名５４１、論理ボリュームブロック番号５４２、記憶装置名５８３、ボリューム名５０１、ボリューム論理ブロック番号５１２、物理記憶装置名５０２、物理ブロック番号５１４の組のデータを保持する。このデータはワーク用空き領域情報６８３は、ホストマッピング情報６５０ｂと記憶装置構成情報１３４を参照して初期化する。このうち、ホスト名６３１とファイルシステムＩＤ５３５に有効値を有するエントリはそれらにより識別されるファイルシステム１０４ｂにおけるデータ移動先となりうる領域を示し、ホストマッピング情報６５０ｂ中のファイル記憶位置情報５３０ｂからその領域を把握する。この領域に関しては、そのファイルシステム１０４ｂ内に存在するデータの移動先として利用可能である。

【０１７６】

ホスト名６３１とファイルシステムＩＤ５３５に無効値を有するがボリューム名５０１に有効値を有するエントリは、まだどのホストからも利用されていない記憶装置１０が提供しているボリューム内の記憶領域を示し、記憶装置構成情報１３４から把握される記憶装置１０が提供している全領域からホストマッピング情報６５０ｂから把握される利用領域を除いた領域として把握する。この領域に関しては、ボリュームマネージャ１０２ｂにおけるデータ移動先として利用可能である。ボリューム名５０１に無効値を有するエントリは記憶装置１０内のデータ移動先として利用可能な領域を示し、記憶装置構成情報１３４から把握可能である。

【０１７７】

データ構造仮想ボリューム内位置情報６９０ｂはＤＢＭＳ１１０が保持しているデータ構造が記憶装置１０が提供するボリュームのどこに記憶されているかを示した情報であり、ホスト名６５１、ＤＢＭＳ名６３１、データ構造名５６１、データファイルパス名５６２、ファイルシステムＩＤ５３５、ファイルブロック番号５６３、記憶装置名５８３、ボリューム名５０１、ボリューム論理ブロック番号５１２の組を保持する。この情報は、ＤＢＭＳスキーマ情報１３６中のＤＢＭＳデータ記憶位置情報６２２とＤＢＭＳホスト情報６２６とホストマッピング情報６５０ｂを参照し、ファイル（ローデバイス）パス、（論理

）ボリュームに関して対応する部分をまとめることにより初期データを作成する。

【０１７８】

図２９に移動プラン情報７５０ｂを示す。移動プラン情報７５０ｂは、ボリュームマネージャ１０２ｂに対するデータ移動指示を記憶する論理ボリューム移動プラン情報７５３とファイルシステム１０４ｂに対するデータ移動指示を記憶するファイルブロック移動プラン情報７５４と記憶装置１０に対するデータ移動指示を記憶する物理記憶位置移動プラン情報７５２を含む。これら情報に関しては、何もデータを持たないように初期化する。図１５の移動プラン情報７５０と比較して、移動プラン情報７５０ｂは仮想ボリューム移動プラン情報７５２が削除され、論理ボリューム移動プラン情報７５３とファイルブロック移動プラン情報７５４が追加されたものとなっている。

【０１７９】

論理ボリューム移動プラン情報７５３には、移動順序７６１、移動処理を行うホスト名６３１、移動元ボリュームとその領域を指定する移動論理ボリューム名７７３、移動論理ボリュームブロック番号７７４、移動先の記憶装置１０とその記憶領域を指定する移動先記憶装置名７６４、移動先ボリューム名７６５、移動先ブロック番号７６６の組が記憶される。ファイルブロック移動プラン情報７５４には、移動順序７６１、移動処理を行うホスト名６５１、移動元ファイルとその領域を指定するファイルシステムＩＤ５３５、移動データファイルパス名７７５、移動ファイルブロック番号７７６、移動先領域を指定する移動先記憶装置名７６４、移動先ボリューム名７６５、移動先ブロック番号７６６の組が記憶される。

＜第三の実施の形態＞

本実施形態では、ＤＢＭＳが実行される計算機とファイルを管理単位とする記憶装置がネットワークを用いて接続された計算機システムにおいて、データの記憶位置の管理を行う計算機が存在し、そこで計算機システム内のデータの記憶位置の管理を行う。計算機上で実行されるＯＳ中のネットワークファイルシステムは、複数のファイルを仮想的な１つのファイルにまとめ、その構成を動的に変更する機能を有する。また、記憶装置においても、記憶装置内部でデータの記憶位置を動的に変更する機能を有する。

【０１８０】

データ記憶位置管理を実施する計算機は、ＤＢＭＳに関する情報、データの記憶位置のマッピングに関する情報、記憶装置の構成情報を取得し、それらを用いて好適なデータ再配置案を作成する。ネットワークファイルシステム，記憶装置に対して作成したデータ配置を指示し、それらのデータ再配置機能を用いてそのデータ再配置案に従ったデータ配置を実現する。

【０１８１】

図３０は、本発明の第三の実施の形態における計算機システムの構成図である。図示されたように、本発明の第三の実施の形態は本発明の第一の実施の形態と以下の点が異なる。

【０１８２】

本実施の形態においてはＩ／Ｏパスインターフェイス７０、Ｉ／Ｏパス７１、仮想ボリュームスイッチ７２が存在せず、記憶制御装置１０ｃとＤＢホスト８０ｃはネットワーク７９を介してのみ接続される。記憶装置１０はファイルを単位としたデータ記憶管理を行う記憶装置１０ｃに変更される。そのため、物理記憶装置稼動情報３２、ボリューム物理記憶位置管理情報３６がそれぞれ物理記憶装置稼動情報３２ｃ、ファイル記憶管理情報３６ｃに変更される。

【０１８３】

ＤＢホスト８０ｃで実行されるＯＳ１００ではボリュームマネージャ１０２、ファイルシステム１０４が削除されその代わりにネットワークファイルシステム１０４ｃが追加され、ＯＳ１００が保持するマッピング情報１０６がマッピング情報１０６ｃへ変更される。データ位置管理サーバ８２内に記憶される記憶装置稼動情報１３２、記憶装置構成情報１３４、データ記憶位置管理情報１４０がそれぞれ記憶装置稼動情報１３２ｃ、記憶装置構成情報１３４ｃ、データ記憶位置管理情報１４０内のホストマッピング情報６５０の内容

を若干変更したホストマッピング情報６５０ｃに変更される。

【０１８４】

記憶装置１０はファイルを管理単位とする記憶装置１０ｃに変更される。ＤＢホスト８０ｃからのアクセスもＮＦＳ等のファイルをベースとしたプロトコルで実施される。記憶装置１０におけるボリュームの役割は、記憶装置１０ｃにおいてはファイルもしくはファイルを管理するファイルシステムとなり、そのファイルの記憶位置管理情報がファイル記憶管理情報３６ｃである。１つの記憶装置１０ｃの中に複数のファイルシステムが存在しても構わない。

【０１８５】

物理記憶装置１８の稼動情報はボリュームを単位とした取得からファイルシステム又はファイルを単位とした取得に変更する。記憶装置１０ｃ内にファイルシステムが存在する場合でもデータの移動機能を実現可能であり、データ移動指示方法は、前と同じく移動するファイルとそのデータ領域と、移動先領域を指定するものとする。本実施の形態においては、記憶装置１０ｃにおけるデータ移動機能は必須であるとする。

【０１８６】

ネットワークファイルシステム１０４ｃは、記憶装置１０ｃが提供するファイルをアクセスするための機能を提供する。更に、複数のファイルを仮想的な１つのファイルとして提供する。この機能を実現するためにネットワークファイルシステム１０４ｃは管理情報をマッピング情報１０６ｃ内に保持し、仮想ファイルアクセス時にこの管理情報を参照し、実際のアクセス先を求める処理を行う。更に、その構成を動的に変更する機能を有する。これら処理は、ＤＢホスト８０上で管理コマンドを実行することにより実施される。

【０１８７】

図３１は記憶装置１０ｃ内に保持される物理記憶装置稼動情報３２ｃを示す。図２の物理記憶装置稼動情報３２からの変更点は、稼動情報取得単位がボリュームからファイルシステムに変更されたため、ボリューム名５０１の部分がファイルシステム名１００１に変更されたことである。また、稼動情報取得単位をファイルとしてもよく、このときはボリューム名５０１の部分がファイルシステム名１００１とファイルパス名１００２に変更される。

図３２は記憶装置１０ｃ内に保持されるファイル記憶管理情報３６ｃを示す。図３のボリューム物理記憶位置管理情報３６からの変更点は、ボリューム物理記憶位置メイン情報５１０、ボリュームデータ移動管理情報５１１からファイル物理記憶位置情報５１０ｃ、ファイルデータ移動管理情報５１１ｃにそれぞれ変更される。上記の変更内容は、ボリュームの識別子がボリューム名５０１がファイルの識別子となるファイルシステム名１００１とファイルパス名１００２に、ボリューム内のデータ領域を示すボリューム論理ブロック番号５１２と移動論理ブロック番号７８２がそれぞれファイルブロック番号１００３または移動ファイルブロック番号１０２１に変更されたことである。

【０１８８】

ここで、ファイルパス名１００２が“Ｅｍｐｔｙ”であるエントリ１０１５は特殊なエントリであり、このエントリには記憶装置１０ｃ内の物理記憶装置１８の領域のうち、指定ファイルシステム内でファイルの記憶領域としてに割り当てられていない領域を示し、図３中のボリュームデータ移動管理情報５１１を用いるデータ移動方式で説明した処理手順を用い、この領域に対して移動するデータをコピーすることによりデータの物理記憶位置の動的変更機能を実現する。

【０１８９】

図３３はＤＢホスト８０ｃのＯＳ１００内に記憶されているマッピング情報１０６ｃを示す。マッピング情報１０６ｃ中には、ネットワークファイルシステムマウント情報１０３０と仮想ファイル情報１０４０と仮想ファイルデータ移動管理情報１０５０が含まれる。

【０１９０】

ネットワークファイルシステムマウント情報１０３０は、記憶装置１０ｃから提供され、ＤＢホスト８０ｃにおいてマウントされているファイルシステムの情報で、ファイルシス

テムの提供元記憶装置とそのファイルシステムの識別子である記憶装置名５８３とファイルシステム名１００１、そして、そのファイルシステムのマウントポイントの情報であるマウントポイント１０３１の組を保持する。

【０１９１】

仮想ファイル情報１０４０は、ネットワークファイルシステム１０４ｃが提供する複数の記憶装置から提供されるファイルを仮想的な１つのファイルとして提供する機能の管理に用いる情報で、提供される仮想ファイルの識別子である仮想ファイルパス名１０４１とそのデータ領域を示す仮想ファイルブロック番号１０４２とそのデータ領域のデータを実際に保持するファイルの識別子である構成ファイルパス名１０４３とその記憶領域を示す構成ファイルブロック番号１０４４の組を含む。

【０１９２】

仮想ファイルデータ移動管理情報は１０５０は、ネットワークファイルシステム１０４ｃが提供する仮想ファイルの構成の変更処理の一部である、構成データの移動処理を行う際に利用する管理情報で、データ移動を行う移動元の仮想ファイルの識別子である移動仮想ファイルパス名１０５１とデータ移動を行う領域を示す移動仮想ファイルブロック番号１０５２とそのデータの移動先ファイルとその移動先領域を示す移動先構成ファイルパス名１０５３と移動先ファイルブロック番号１０５４、そして、データ移動処理を行う際の管理情報である差分管理情報７８５とコピーポインタ７８６の組を含む。データの移動先に関しては、移動先ファイルの移動先指定領域に記憶領域実体が確保されている以外の制約は存在しない。本情報を利用し、図３中のボリュームデータ移動管理情報５１１を用いるデータ移動方式で説明した処理手順を用いることにより、データの移動機能を実現できる。

【０１９３】

図３４はデータ位置管理サーバ８２ｃ上に記憶される記憶装置稼動情報１３２ｃを示す。図８の記憶装置稼動情報１３２からの変更点は、記憶装置稼動情報１３２ではボリューム単位で稼動情報を取得していたものを記憶装置稼動情報１３２ｃではファイルシステム単位で稼動情報を取得していることである。そのため、ボリューム名５０１がファイルシステム名１００１に変更される。

【０１９４】

図３５はデータ位置管理サーバ８２ｃ上に記憶される記憶装置構成情報１３４ｃを示す。図９の記憶装置構成情報１３４からの変更点は、記憶装置毎のボリューム物理記憶位置メイン情報５１０を記憶する記憶装置ボリューム物理記憶位置管理情報６０４が記憶装置毎のファイル物理記憶位置情報５１０ｃを記憶する記憶装置ファイル物理記憶位置情報６０４ｃに変更される。

【０１９５】

図３６はデータ位置管理サーバ８２ｃ上に記憶されるホストマッピング情報６５０ｃを示す。図１２のホストマッピング情報６５０からの変更点は、各ホストのマッピング情報１０６を保持するマッピング情報６５２が各ホストのマッピング情報１０６ｃを保持するマッピング情報６５２ｃになったことである。

【０１９６】

データ再配置処理において、データ移動機能を保持する部分が仮想ボリュームスイッチ７２からネットワークファイルシステム１０４ｃに変更され、記憶装置１０ｃがファイルを管理単位とするように変更されたことによる変更点は以下のようなものである。

【０１９７】

ネットワークファイルシステム１０４ｃにおいては、ＤＢホスト８０ｃ上において管理コマンドを実施することによりデータの移動を実施する。そこで、データ位置管理主プログラム１３０がネットワーク７９を通して、ＤＢホスト８０ｃ上で実行されているデータ位置管理副プログラム１２０に、ネットワークファイルシステム１０４ｃにおいてデータの移動を実施する管理コマンドを実行する指示を出し、それに従ってデータ位置管理副プログラム１２０が管理コマンドを実施することによりデータの移動を実施する。このとき、

現在空きの領域に対するデータの移動を行うため、データ移動先領域となるファイルやファイル内の領域が存在しない。このようなデータ移動指示を受け取った場合には、ネットワークファイルシステム１０４ｃは指定ファイルの新規ファイル作成や領域拡張を実施し、それが成功した後にデータ移動処理を開始し、データ移動中の領域不足の問題を回避する。

【０１９８】

また、ネットワークファイルシステム１０４ｃは通常のプロトコルを利用してファイル作成や領域拡張を実施する。そのため、そのデータ記憶先が必ずしも最適な場所に割り当てられるとは限らない。そこで、ネットワークファイルシステム１０４ｃによるデータ移動が完了後、今度は記憶装置１０ｃに対して記憶装置内データ移動の指示を出し、作成したデータ再配置案に従ったデータ配置を実現する。このとき、記憶装置１０ｃ内でデータ移動元と移動先が重なっている場合には、一旦データを移動先とは異なる空き領域にデータを移動し、その後に再度指定された移動先にデータを移動させる処理を記憶装置１０ｃは実行する。

【０１９９】

データ再配置案作成処理においては、データ再配置案を作成する際に利用するワーク領域であるデータ再配置ワーク情報６７０がデータ再配置ワーク情報６７０ｃに、移動プラン情報７５０が移動プラン情報７５０ｃに変更される。

【０２００】

図３７にデータ再配置ワーク情報６７０ｃを示す。データ再配置ワーク情報６７０ｃ中には、記憶装置ファイル物理記憶位置情報６８１ｃとデータ構造記憶装置内ファイル位置情報６９０ｃが含まれる。図１４のデータ再配置ワーク情報６７０と比べて、仮想ボリューム物理記憶位置情報６８０は記憶装置ファイル物理記憶位置情報６８１ｃに変更され、データ構造仮想ボリューム内位置情報６９０はデータ構造仮想ボリューム内位置情報６９０ｃに変更される。

【０２０１】

記憶装置ファイル物理記憶位置情報６８１ｃは記憶装置１０ｃが提供するファイルシステムとその内部に存在するファイル、その物理記憶装置１８における記憶位置の一覧であり、記憶装置名５８３、ファイルシステム名１００１、ファイルパス名１００２、ファイルブロック番号１００３、物理記憶装置名５０２、物理ブロック番号５１４の組を保持する。このデータは、記憶装置構成情報１３４ｃを参照して初期化する。ファイルパス名１００２が“Ｅｍｐｔｙ”であるエントリ１０７１は記憶装置１０ｃのファイルシステム名１００１の領域のうち、ファイルの記憶に利用されていない領域を示し、ここに対してデータの移動が可能である。

【０２０２】

データ構造仮想ボリューム内位置情報６９０ｃはＤＢＭＳ１１０が保持しているデータ構造が記憶装置１０ｃが提供するファイルのどこに記憶されているかを示した情報であり、ホスト名６５１、ＤＢＭＳ名６３１、データ構造名５６１、データファイルパス名５６２、ファイルブロック番号５６３、記憶装置名５８３、ファイルシステム名１００１、ファイルパス名１００２、ファイルブロック番号１００３の組を保持する。この情報は、ＤＢＭＳスキーマ情報１３６中のＤＢＭＳデータ記憶位置情報６２２とＤＢＭＳホスト情報６２６とホストマッピング情報６５０ｃを参照し、ファイルパスに関して対応する部分をまとめることにより初期データを作成する。

【０２０３】

図３８に移動プラン情報７５０ｃを示す。移動プラン情報７５０ｃは、ネットワークファイルシステム１０４ｃに対するデータ移動指示を記憶する仮想ファイルブロック移動プラン情報７５５と記憶装置１０ｃに対するデータ移動指示を記憶する物理記憶位置移動プラン情報７５２ｃを含む。これら情報に関しては、何もデータを持たないように初期化する。図１５の移動プラン情報７５０と比較して、移動プラン情報７５０ｃは仮想ボリューム移動プラン情報７５２が削除され、仮想ファイル移動プラン情報７５５が追加され、物理

記憶位置移動プラン情報７５２が物理記憶位置移動プラン情報７５２ｃに変更されたものとなる。

【０２０４】

仮想ファイル移動プラン情報７５５には、移動順序７６１、移動処理を行うホスト名６５１、移動元の仮想ファイルとその領域を指定する移動仮想ファイルパス名１０５１、移動仮想ファイルブロック番号１０５２、移動先の構成ファイルとその領域を指定する移動先構成ファイルパス名１０５３、移動先ファイルブロック番号１０５４の組が記憶される。物理記憶位置移動プラン情報７５２ｃには、移動順序７６１、移動処理を行う移動記憶装置名７６７、移動元のファイルとその領域を指定する移動ファイルシステム名１１０１、移動ファイルパス名１１０２、移動ファイルブロック番号１１０３、移動先の物理記憶装置１８とその領域を指定する移動先物理記憶装置名７７１、移動先物理ブロック番号７７２の組が記憶される。

【０２０５】

本実施例においては、ネットワークファイルシステム１０４ｃが複数のファイルから１つの仮想ファイルを構成する機能を有しているが、ここで、単純に１つの仮想ファイルは１つのファイルから構成されても構わない。このとき、ネットワークファイルシステム１０４ｃは動的なデータ移動機能のみを提供する。さらに、ＤＢＭＳ１１０が処理を中断しても構わない場合には、ＤＢＭＳ１１０の処理の中断後、ファイルのコピーを実施し、アクセス先がコピー先になるようにシンボリックリンクを張りなおした後にＤＢＭＳ１１０が処理を再開することによるデータ再配置も可能である。さらに、ネットワークファイルシステム１０４ｃにおけるデータ移動を行なわず、記憶装置１０ｃのみでデータ移動を実施することも可能である。

【０２０６】

【発明の効果】

本発明により以下のことが可能となる。第一に、ＤＢＭＳが管理するデータの特性を考慮して記憶装置におけるデータ記憶位置を決定することにより、ＤＢＭＳに対してより好ましいアクセス性能特性を持つ記憶装置を保持する計算機システムが実現される。これにより、その計算機システムで稼動しているＤＢＭＳの性能を向上させることができる。特に、複数の記憶装置を利用するＤＢシステムにおいて、各記憶装置間へアクセス要求が適切に分散化されるようにし、ＤＢＭＳの処理性能を向上させる。

【０２０７】

第二に、ＤＢＭＳが稼動している計算機システムにおいて、ＤＢＭＳの特性を考慮した記憶装置の良好なアクセス性能特性達成を目的としたデータ記憶位置再配置処理を実現するため、計算機システムの性能に関する管理コストを削減することができる。特に、本発明を用いることにより、データ記憶位置の再配置案を自動で作成することが可能であり、管理コストの削減に大きく寄与する。更に、多数のＤＢＭＳが稼動し、多数の記憶装置が存在するシステムにおいても利用可能で集中的な管理を可能とするため、そのようなシステムにおける性能に関する管理コストを削減することができる。

【図面の簡単な説明】

【図１】第一の実施の形態における計算機システムの構成を示す図である。

【図２】記憶装置１０内に保持されている物理記憶装置稼動情報３２を示す図である。

【図３】記憶装置１０内に保持されているボリューム物理記憶位置管理情報３６を示す図である。

【図４】第一の実施の形態におけるＤＢホスト８０のＯＳ１００内に記憶されているマッピング情報１０６を示す図である。

【図５】ＤＢＭＳ１１０内に記憶されているその内部で定義・管理しているデータその他の管理情報であるスキーマ情報１１４を示す図である。

【図６】ＤＢホスト８０のメモリ８８上に記憶されている実行履歴情報１２２を示す図である。

【図７】仮想ボリュームスイッチ７２が保持する仮想ボリューム情報７３を示す図である

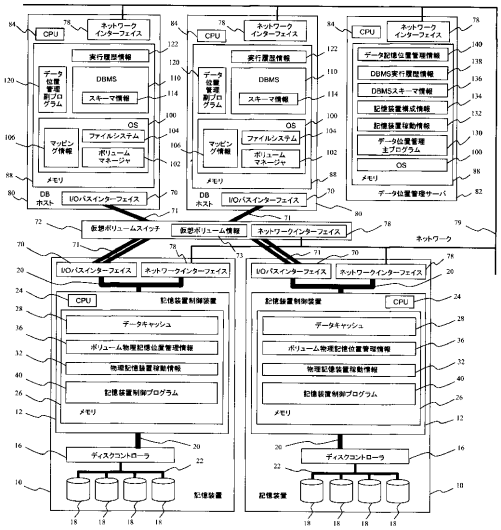
。
【図８】データ位置管理サーバ８２上に記憶される記憶装置稼動情報１３２を示す図である。
【図９】データ位置管理サーバ８２上に記憶される記憶装置構成情報１３４を示す図である。
【図１０】データ位置管理サーバ８２上に記憶されるＤＢＭＳスキーマ情報１３６を示す図である。
【図１１】データ位置管理サーバ８２上に記憶されるＤＢＭＳ実行履歴情報１３８を示す図である。
【図１２】データ位置管理サーバ８２上に記憶されるデータ記憶位置管理情報１４０を示す図である。
【図１３】データ位置管理主プログラム１３０によるデータ再配置処理の処理フローを示す図である。
【図１４】データ配置解析・再配置案作成処理で利用するデータ再配置ワーク情報６７０を示す図である。
【図１５】データ配置解析・再配置案作成処理により作成されるデータ移動案を格納する移動プラン情報７５０を示す図である。
【図１６】記憶装置稼動情報１３２を基にした同時アクセス実行データ構造を分離するためのデータ再配置案作成処理の処理フローを示す図である。
【図１７】ＤＢＭＳ実行履歴情報１３８を基にした同時アクセス実行データ構造を分離するためのデータ再配置案作成処理で利用するクエリ実行時同時アクセスデータカウント情報７００を示す図である。
【図１８】ＤＢＭＳ実行履歴情報１３８を基にした同時アクセス実行データ構造を分離するためのデータ再配置案作成処理処理フローを示す図である。
【図１９】指定されたデータ構造とそのデータ構造と同時にアクセスされる可能性が高いデータ構造の組を分離するデータ再配置案を作成する処理のフローを示す図である。
【図２０】データ構造の定義を基にした同時アクセス実行データ構造を分離するためのデータ再配置案作成処理の処理フローを示す図である。
【図２１】特定の表や索引の同一データ構造に対するアクセス並列度を考慮したデータ再配置案作成処理の処理フローを示す図である。
【図２２】特定の表データに対するシーケンシャルアクセス時のディスクネックを解消するデータ再配置案作成処理の処理フローを示す図である。
【図２３】特定のデータ構造に対する記憶装置１０におけるキャッシュ効果を考慮したデータ再配置案作成処理で利用するキャッシュ効果判定処理の処理フローを示す図である。
【図２４】特定のデータ構造に対する記憶装置１０におけるキャッシュ効果を考慮したデータ再配置案作成処理の処理フローを示す図である。
【図２５】第二の実施の形態における計算機システムの構成を示す図である。
【図２６】ＤＢホスト８０ｂのＯＳ１００内に記憶されるマッピング情報１０６ｂを示す。
【図２７】データ位置管理サーバ８２上に記憶されるホストマッピング情報６５０ｂを示す図である。
【図２８】データ配置解析・再配置案作成処理で利用するデータ再配置ワーク情報６７０ｂを示す図である。
【図２９】データ配置解析・再配置案作成処理により作成されるデータ移動案を格納する移動プラン情報７５０ｂを示す図である。
【図３０】第三の実施の形態における計算機システムの構成を示す図である。
【図３１】記憶装置１０ｃ内に保持される物理記憶装置稼動情報３２ｃを示す図である。
【図３２】記憶装置１０ｃ内に保持されるファイル記憶管理情報３６ｃを示す図である。
【図３３】ＤＢホスト８０ｃのＯＳ１００内に記憶されているマッピング情報１０６ｃを示す図である。

【図３４】データ位置管理サーバ８２ｃ上に記憶される記憶装置稼動情報１３２ｃを示す図である。
【図３５】データ位置管理サーバ８２ｃ上に記憶される記憶装置構成情報１３４ｃを示す図である。
【図３６】データ位置管理サーバ８２ｃ上に記憶されるホストマッピング情報６５０ｃを示す図である。
【図３７】データ配置解析・再配置案作成処理で利用するデータ再配置ワーク情報６７０ｃを示す図である。
【図３８】データ配置解析・再配置案作成処理で利用する移動プラン情報７５０ｃを示す図である。

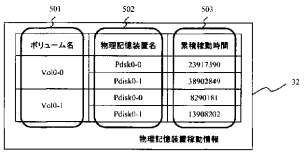
【符号の説明】

１０，１０ｃ	記憶装置
１８	物理記憶装置
２８	データキャッシュ
３２，３２ｃ	物理記憶装置稼動情報
３６	ボリューム物理記憶位置管理情報
３６ｃ	ファイル記憶管理情報
７０	Ｉ／Ｏパスインターフェイス
７１	Ｉ／Ｏパス
７２	仮想ボリュームスイッチ
７２ｂ	Ｉ／Ｏパススイッチ
７３	仮想ボリューム情報
７８	ネットワークインターフェイス
７９	ネットワーク
８０，８０ｂ，８０ｃ	ＤＢホスト
８２	データ位置管理サーバ
１００	ＯＳ（オペレーティングシステム）
１０２，１０２ｂ	ボリュームマネージャ
１０４，１０４ｂ	ファイルシステム
１０４ｃ	ネットワークファイルシステム
１０６，１０６ｂ，１０６ｃ	マッピング情報
１１０	ＤＢＭＳ（データベース管理システム）
１１４	スキーマ情報
１２０	データ位置管理副プログラム
１２２	実行履歴情報
１３０	データ位置管理主プログラム
１３２，１３２ｃ	記憶装置稼動情報
１３４，１３４ｃ	記憶装置構成情報
１３６	ＤＢＭＳスキーマ情報
１３８	ＤＢＭＳ実行履歴情報
１４０	データ記憶位置管理情報
６５０，６５０ｂ，６５０ｃ	ホストマッピング情報

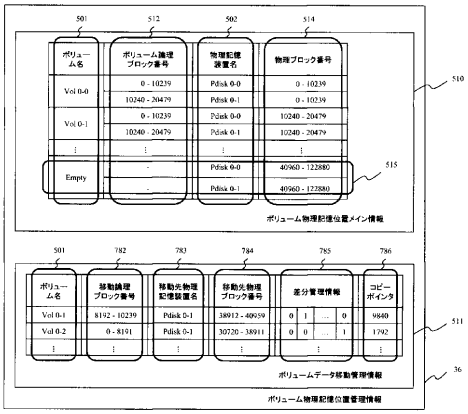
【図１】



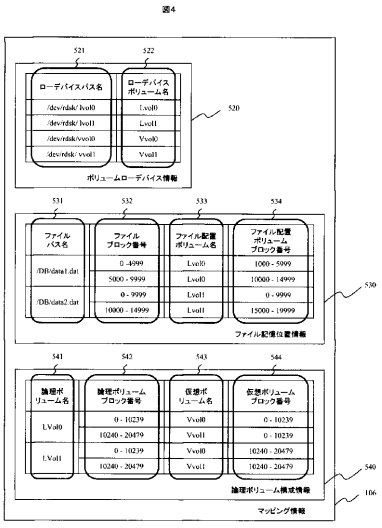
【図２】



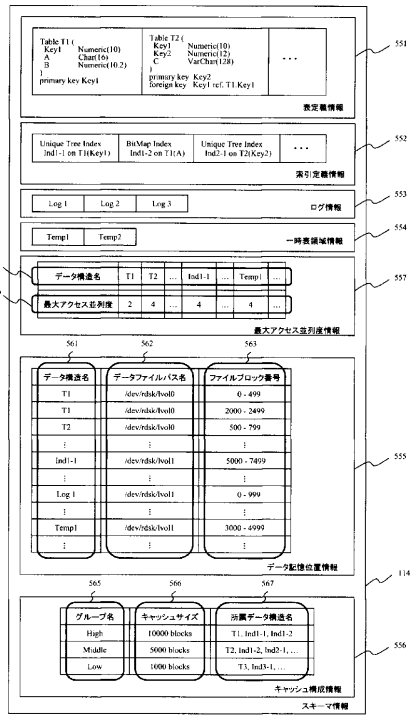
【図３】



【図４】

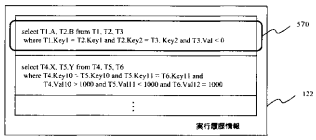


【図５】



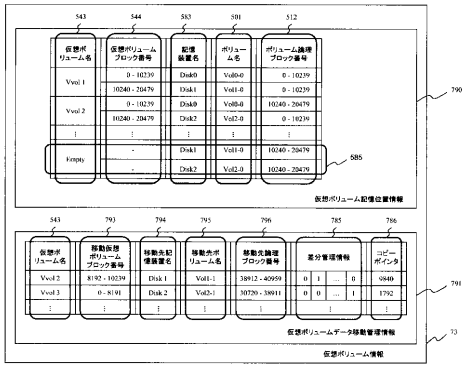
【図６】

図6



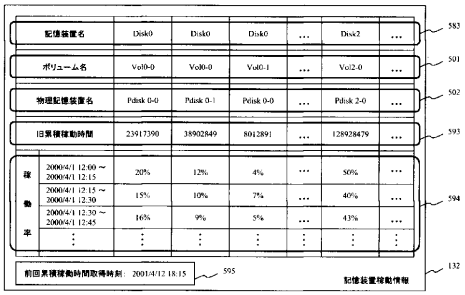
【図７】

図7



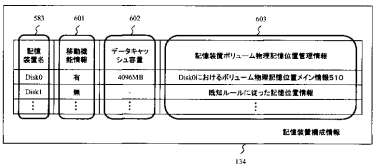
【図８】

図8



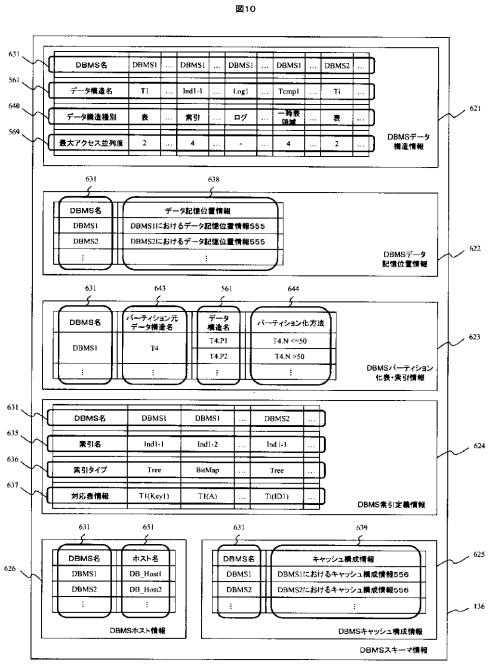
【図９】

図9



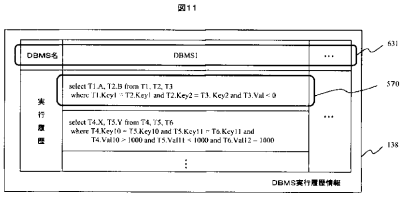
【図１０】

図10



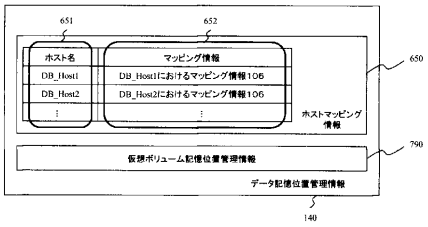
【図１１】

図11



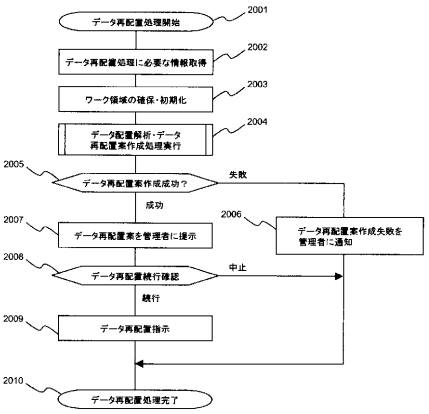
【図１２】

図12



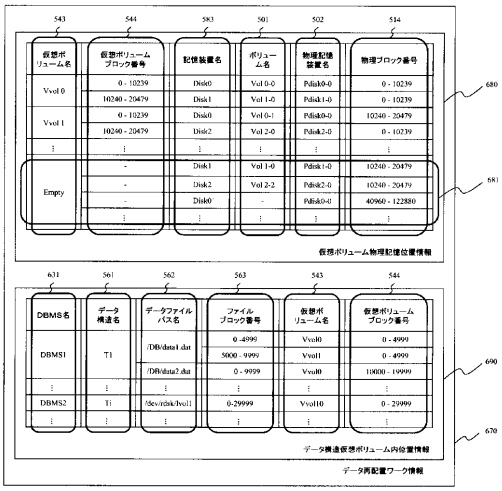
【図１３】

図13



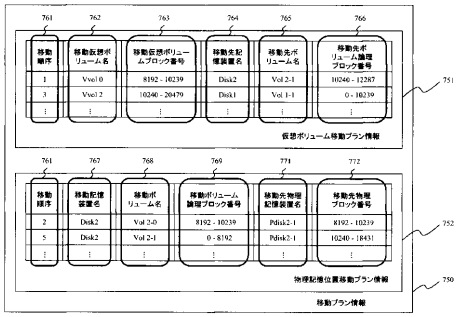
【図１４】

図14



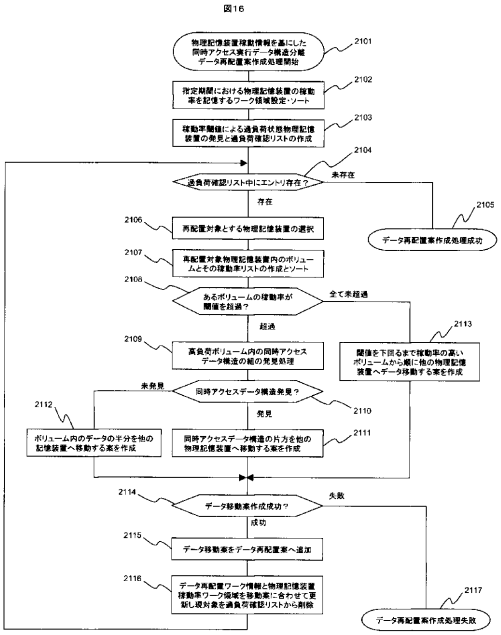
【図１５】

図15



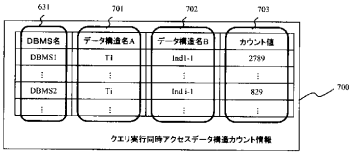
【図１６】

図16



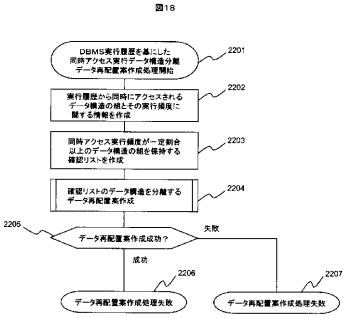
【図１７】

図17



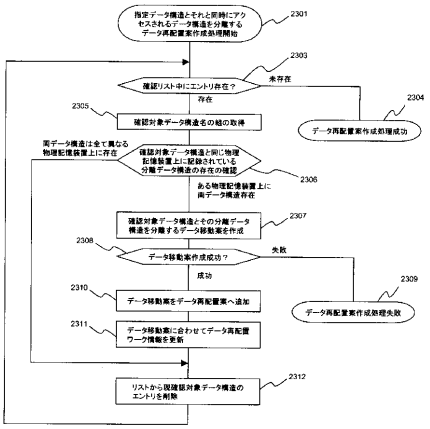
【図１８】

図18



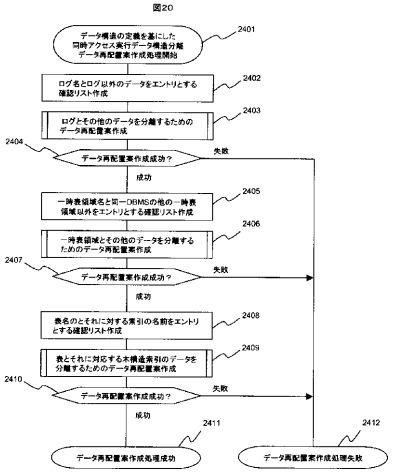
【図１９】

図19



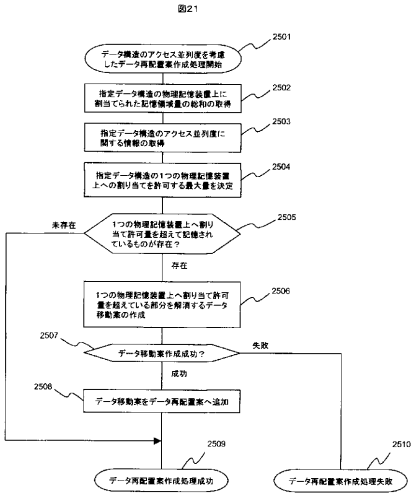
【図２０】

図20



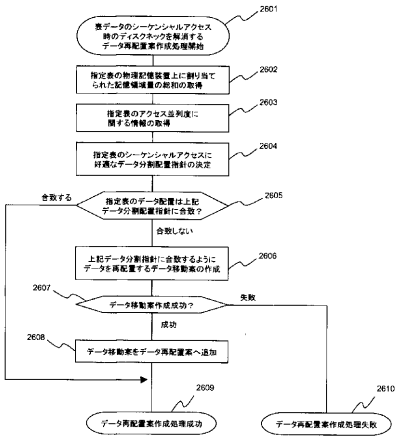
【図２１】

図21



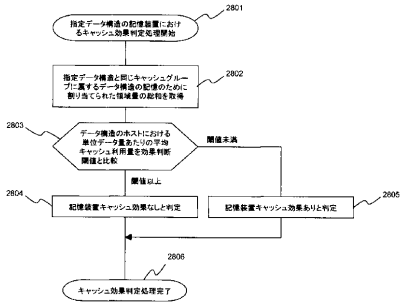
【図２２】

図22



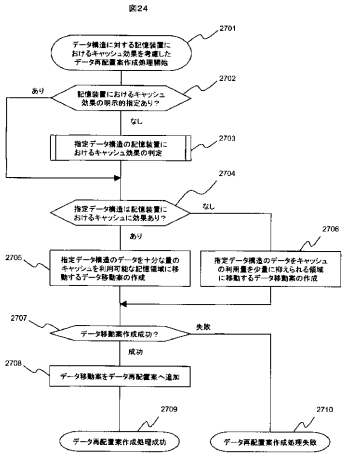
【図２３】

図23



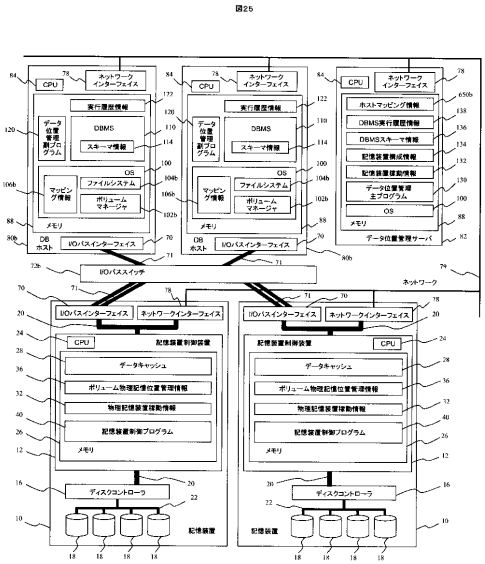
【図２４】

図24



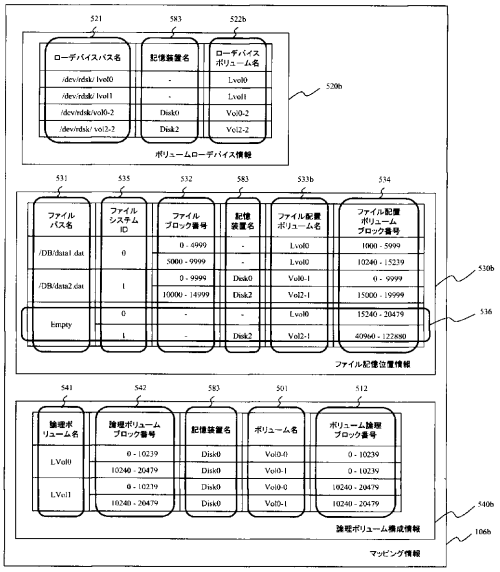
【図２５】

図25



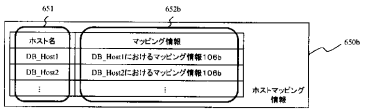
【図２６】

図26



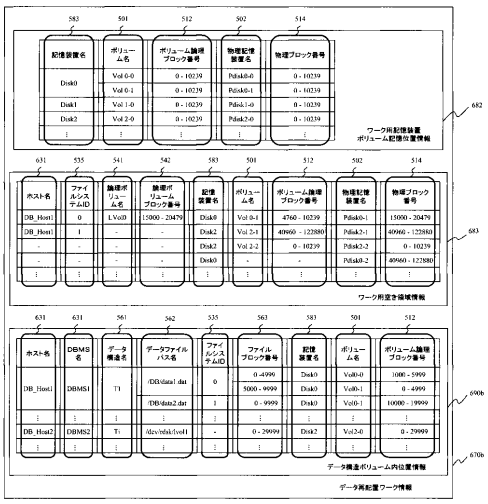
【図２７】

図27



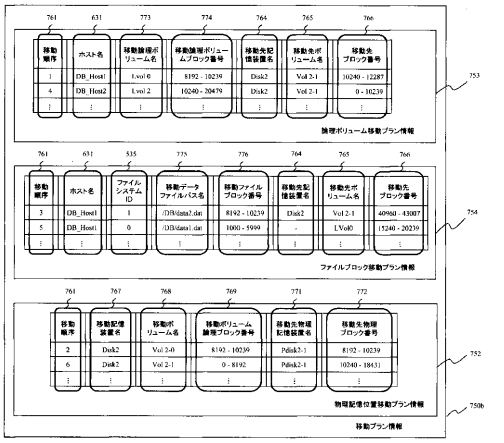
【図２８】

図28



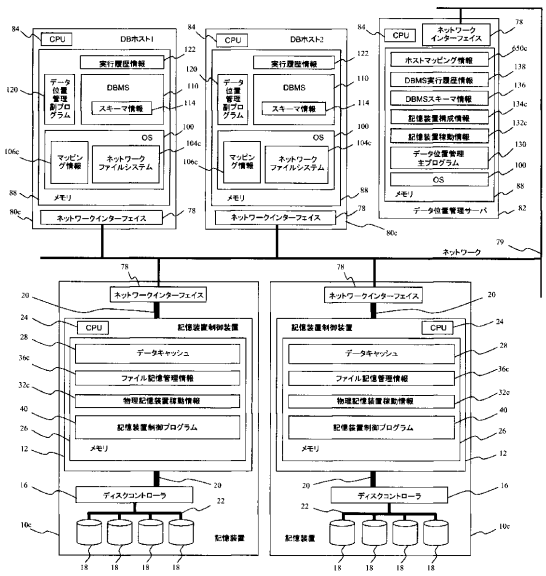
【図２９】

図29



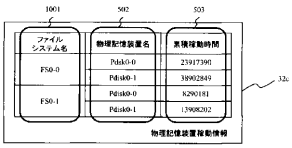
【図３０】

図30



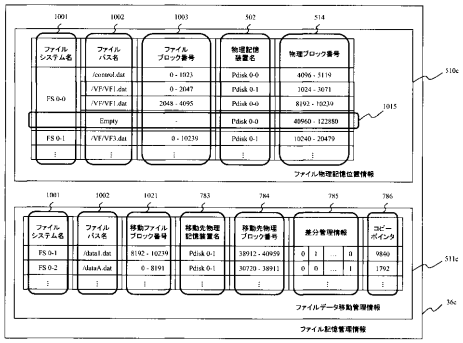
【図３１】

図31



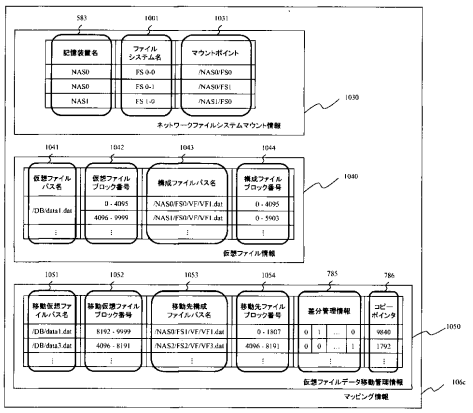
【図３２】

図32



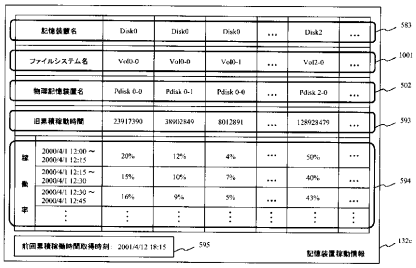
【図３３】

図33



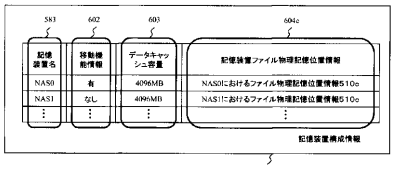
【図３４】

図34



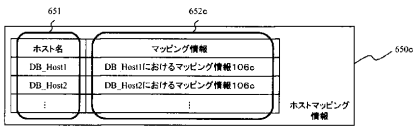
【図３５】

図35



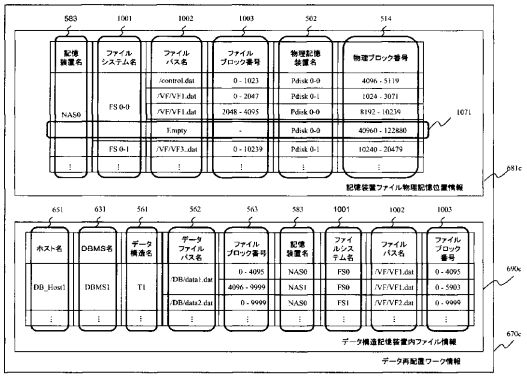
【図３６】

図36



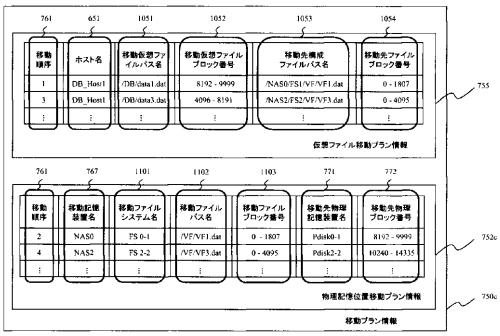
【図３７】

図37



【図３８】

図38


761
651
1051
1052
1053
1054
移動順序
ホスト名
移動仮想ファイルパス名
移動仮想ファイルブロック番号
移動先構成ファイルパス名
移動先ファイルブロック番号
1
DB_Host1
/DB/data1.dat
8192 - 9999
/NAS0/FS1/VF/VF1.dat
0 - 1807
3
DB_Host1
/DB/data3.dat
4096 - 8191
/NAS2/FS2/VF/VF3.dat
0 - 4095
755
仮想ファイル移動プラン情報
761
767
1101
1102
1103
771
772
移動順序
移動記憶装置名
移動ファイルシステム名
移動ファイルパス名
移動ファイルブロック番号
移動先物理記憶装置名
移動先物理ブロック番号
2
NAS0
FS 0-1
/VF/VF1.dat
0 - 1807
Pdisk0-1
8192 - 9999
4
NAS2
FS 2-2
/VF/VF3.dat
0 - 4095
Pdisk2-2
10240 - 14335
752c
物理記憶位置移動プラン情報
750c
移動プラン情報

フロントページの続き

(72)発明者　喜連川　優
千葉県松戸市二十世紀が丘丸山町１７

審査官　高瀬　勤

(56)参考文献　特開平０５－０１２３３８（ＪＰ，Ａ）
原　憲宏 他，並列ＤＢサーバシステムにおけるＤＢ再配置方式の提案，第４８回（平成６年前期）全国大会講演論文集（４），日本，社団法人情報処理学会，１９９４年　3月２５日，p.207-208

(58)調査した分野(Int.Cl.，ＤＢ名)
G06F　12/00
G06F　17/30
JSTPlus(JDream2)